ANGLIA RUSKIN UNIVERSITY

THE EFFECT OF DYNAMIC RANGE COMPRESSION ON THE PSYCHOACOUSTIC QUALITY AND LOUDNESS OF COMMERCIAL MUSIC

WILLIAM MARK CAMPBELL

A thesis in partial fulfilment of the requirements of Anglia Ruskin University for the degree of Doctor of Philosophy

Submitted: December 2018

Acknowledgements

Thank you, Dr Ian van der Linde, for all the unmeasurable time you have spent helping me, beyond the requirements of a second supervisor. I would also like to credit Dr van der Linde with the MATLAB programming aspects (with my assistance) of this research, for which I am truly grateful.

Thank you, Dr Ryan Stables, for the wakeup call after four years of being adrift and for pointing out that for my work to be taken seriously I must adapt to the 'correct' method of researching my subject.

Thank you, Professor Justin Paterson for contending with my lack of writing expertise and being a shoulder to cry on over the years.

Thank you, Dr Douglas Nunn, for all your proofreading skills and your attention to detail concerning aspects of my research.

Thank you, Dr Robert Toulson for asking me to take on the 'Loudness War' research; it is a subject near and dear to me, and I have thoroughly enjoyed the work.

Thank you to the bands and musicians whom I have recorded and allowed me to use their stems for test data - Bijoumiyo, Echo Trails, and State Villains. Special thanks to Tony Rhone for the recording of guitar signals at short notice.

Thank you to my children Toni and Mark for persevering, despite all we have been through during this long process. You were teenagers when I started, and now you are both young adults whom I am very proud of, and I love you both very much.

Thank you, to the anonymous JAES article reviewer who was critical of our submission. The meticulous attention to detail you provided demonstrated the standard to work toward in all my writing and especially this thesis.

Finally, I'd like to thank the Faculty of Science and Engineering, Anglia Ruskin University for funding my doctoral research, both as a full-time member of academic staff and previously as a graduate teaching associate.

ANGLIA RUSKIN UNIVERSITY

ABSTRACT

FACULTY OF SCIENCE AND ENGINEERING

DOCTOR OF PHILOSOPHY

THE EFFECT OF DYNAMIC RANGE COMPRESSION ON THE PSYCHOACOUSTIC QUALITY AND LOUDNESS OF COMMERCIAL MUSIC

WILLIAM MARK CAMPBELL

Submitted: December 2018

This thesis studied the effects of Dynamic Range Compression (DRC) on audio signals. This research devised and tested production strategies to improve and reduce the impact of DRC on signals, verified by listener preference.

The nonlinear characteristics of DRC, combined with the interaction of signals once summed, are likely to produce Intermodulation Distortion (IMD), which is unpleasant to hear. In a bid to reduce these nonlinear effects, the point of application, along with the magnitude and type of DRC used in the mixing signal chain was experimented with, reducing the number of signals interacting while under DRC. The different DRC configurations were used to examine fatigue and listener preference.

Listening preference tests from this research demonstrate listener inclination for compression being applied to sums of fewer sources, as opposed to compressing signals formed from many sources or subgroups, as is the traditional method for music production.

Comparative quantitative analysis of simple and compound signal structures under DRC showed some effects from nonlinearity to be the realignment of harmonic signal structures, alteration of instruments' amplitudes relative to one another, reduction of spectral and temporal clarity, and rearranged dynamic variances related to the rhythmic structure of musical signals. This research shows that decreasing the number of signals interacting under DRC, utilising moderate DRC and applying compression rather than limiting type DRC can reduce the effects of intermodulation distortion, and improve listener enjoyment.

Listening tests employing a temporal estimation task showed that heavy DRC signals might induce fatigue, though the results were inconclusive.

KEYWORDS: dynamic range compression, fatigue, listener preference, intermodulation, nonlinear DSP, loudness war, quality, loudness, channel, subgroup, master-buss.

Research Outputs Associated with this Thesis

Campbell, W., 2017. Listener preferences for alternative dynamic-range-compressed audio configurations. *Auditory Perception Group in the Department of Psychology*, University of Cambridge, UK. 17 November 2014. [oral presentation]

Campbell, W., Paterson, J. and van der Linde, I., 2017. Listener preferences for alternative dynamic-range-compressed audio configurations. *Journal of the Audio Engineering Society*, 65(7/8), pp.540-551. [journal article]

Campbell, W., 2014. Strategies for reducing the effects of dynamic range compression. *Faculty of Science & Technology: Pushing the Boundaries – 4th Annual Research and Scholarship Conference*, Anglia Ruskin University, Cambridge, UK. 14 May 2014. [poster]

Campbell, W., Paterson, J. and Toulson, R., 2014. A quantitative evaluation of signal masking in summed and compressed audio. *KES Transactions on Innovation in Music, Special Edition - Innovation in Music 2013.* Shoreham-by-Sea, UK: Future Technology Press, 1(1), pp.20-31. [peer-reviewed abstract, oral presentation delivered by Campbell and full article]

Toulson, E. R., Paterson, J. L. and Campbell, W., 2014. Evaluating harmonic and intermodulation distortion of mixed signals processed with dynamic range compression. *KES Transactions on Innovation in Music, Special Edition - Innovation in Music 2013.* Shoreham-by-Sea, UK: Future Technology Press, 1(1), pp.224–246. [peer-reviewed abstract, oral presentation delivered by Toulson and full article]

Campbell, W., 2012. Compression: Investigations into the Loudness War. *Cultures of the Digital Economy Conference*, Anglia Ruskin University, Cambridge, UK. 27-28 March 2012. [oral presentation]

Ward, J. and Campbell, W., 2010. Loss of our musical heritage: The rise of the digital re-master. *The Art of Record Production Conference*, Leeds Metropolitan

University, Leeds, UK. 3-5 December 2010. [peer-reviewed abstract, oral presentation delivered by Ward]

Campbell, W., Toulson, R. and Paterson, J., 2010. The effect of dynamic range compression on the psychoacoustic quality and loudness of commercial music. *39th International Congress on Noise Control Engineering 2010 (Inter-Noise: Noise and sustainability, 2010),* Lisbon, Portugal, 13-16 June 2010. Red Hook, NY, USA: Curran Associates, Inc. pp.2580-2588. [peer-reviewed abstract, oral presentation delivered by Campbell and full article]

Toulson, E.R. and Campbell, W., 2009. The life and death of dynamic range: who decides how loud? *The Art of Record Production Conference*, Cardiff, Wales, UK, 13-15 November, 2009. [peer-reviewed abstract, oral presentation delivered by Toulson]

Table Of Contents

List of Figures

List of Tables

List of Abbreviations

ABS diff.	Absolute differential
AES	Audio Engineering Society
AN	Auditory Nerve
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ATSC	Advanced Television Systems Committee
BSLoud	Bark Scale Loudness
CD	Compact Disc
CNC	Cochlear Nucleus Complex
CRM	Coordinate Response Measure
dBFS	Decibels Full Scale
DC	Direct Current
DRC	Dynamic Range Compression
EBU	European Broadcast Union
ERB	Equivalent Rectangular Bandwidth
FET	Field Effect Transistor
FFT	Fast Fourier Transform
HC	Heavy Compression
HCF	Heavy Compression Full-Sum
HCS	Heavy Compression Subgroup
HCT	Heavy Compression Track
HF	High Frequency
HL	Heavy Limiting
HLF	Heavy Limit Full-Sum
HLS	Heavy Limit Subgroup
HLSD	High Level Sample Density
HLT	Heavy Limit Track

HRL	Harmonically Related
HUR	Harmonically Unrelated
IHC	Inner Hair Cells
IMD	Intermodulation Distortion
INDEP	Independent Signals
ISO	International Organization for Standardization
ITU	International Telecommunications Union
LEQ	Loudness Equalised
LFC	Log Frequency Centroid
LKFS	Loudness K-Weighted Full Scale
LRA	Loudness Range
LTM	Long-Term Memory
LU	Loudness Units
LUFS	Loudness Units Full Scale
MAE	Mean Absolute Error
MC	Moderate Compression
MCF	Moderate Compression Full-Sum
MCS	Moderate Compression Subgroup
MCT	Moderate Compression Track
ML	Moderate Limiting
MLF	Moderate Limit Full-Sum
MLS	Moderate Limit Subgroup
MLT	Moderate Limit Track
MTE	Mean Time Estimation
MUSHRA	Multi-Stimulus test with Hidden Reference and Anchor
NC	No Compression
OHC	Outer Hair Cells
PPM	Peak Programme Meter

PRRC	Peak to RMS Regression Coefficient
PTS	Permanent Threshold Shift
RDL	Reference Dynamic Level
RLB	Revised Low Frequency, K And B-Weighted Filters
RMS	Root Mean Square
SIR	Spontaneous Impulse Rates
SLM	Sound-Level Meter
SNR	Signal-To-Noise Ratio
SPL	Sound Pressure Level
T_{card}	Time taken to fill the scorecard
TM	Tectorial Membrane
TTS	Temporary Threshold Shift
T_{visdis}	Time taken to clear the visual distractor
UNEQ	Loudness Un-equalised
VI	Volume Indicator
VU	Volume Unit
WM	Working Memory
XMOD	Signals modulating together (cross-modulation)

THE EFFECT OF DYNAMIC RANGE COMPRESSION ON THE PSYCHOACOUSTIC QUALITY AND LOUDNESS OF COMMERCIAL MUSIC

WILLIAM MARK CAMPBELL

DATE: December 2018

Chapter 1:

Introduction

1.1 Context

DRC (dynamic range compression) is a signal manipulation technique used to reduce dynamic variations in audio (Giannoulis, Massberg and Reiss, 2012). Used with gain, dynamically compressed signals exhibit higher Root Mean Square (RMS) amplitude, increasing overall loudness (Moore et al., 2003; Vickers, 2010). In the music industry, this increase in loudness became evident to producers and managers when their artists' music was played alongside music from artists signed to competing record labels on jukeboxes and radio broadcasts, and it came to be believed that increased loudness improved record sales (Packwood, 1974; Schubert, 2004a; Aarseth, 2012; Jones, 2005; Katz, 2007; Brownstein, 2012), leading to the so-called 'loudness war' (Sterne, 2006; Katz, 2007; Moore et al., 2003).

DRC was first used in the late 1920s and early 1930s, and was referred to as *voltage controlled amplification* (VCA) (Giannoulis, Massberg and Reiss, 2012; Plenge, Spikofski and Theile, 1985; Zölzer, 2011). The original purpose of DRC was to reduce the amplitude of an audio signal's inherent dynamic variation (micro-dynamics) to conform to the physical constraints of the intended reproduction medium (Plenge, Spikofski and Theile, 1985; Sterne, 2006; Blesser, 1968); for example optical signals on cellulose film, cut grooves on shellac and later vinyl records, and broadcast radio signals (Bohn, 2011; Plenge, Spikofski and Theile, 1985).

Modern digital media have a larger dynamic capacity than their analogue predecessors (Sterne, 2006; Katz, 2007; Shuttleworth, 2013), yet modern recordings tend to have DRC applied to the extent that the dynamic variations produced are smaller than their analogue counterparts of the late 20th century (Sterne, 2006; Katz, 2007 pp.73; Donahue, 2008), often termed *hyper-compression*. DRC is still required in playback environments in which dynamics are lost to due background noise (Katz, 2007; Sterne, 2006), and for the protection of apparatus such as loudspeakers (Newell, 1999; Katz, 2007; Shuttleworth, 2013). However, current DRC practices far exceed the technical requirements demanded by modern media and audio reproduction apparatus, and even exceed the requirements of earlier analogue technology (Shuttleworth, 2013; Vickers, 2010).

In an interview for Mix magazine (Jones, 2005), Mastering Engineer Stephen Marcussen postulated that one of the main driving factors fuelling hyper-compression is competition between record companies carried over from the era of the 7” 45-rpm vinyl single and jukeboxes. Jukeboxes situated in public meeting places such as diners and dance halls had a fixed playback volume and allowed consumers to choose which tracks were to be played from the catalogue of 7” singles held in the machine (Weiss, 2010). These technologies allowed the direct comparison of competing productions by the consumer (and producers) for the first time. Ultimately, singles that made greater use of DRC were perceived as louder, and record companies assumed the stance that ‘louder is better’ (Packwood, 1974; Schubert, 2004a), pushing intrinsic signal levels up (Aarseth, 2012; Jones, 2005; Katz, 2007; Brownstein, 2012). As an example, 1960s Motown Record Corp. engineer Bob Olhsson, who became a mastering engineer in the 1990s, revealed:

> “*At Motown [records], we took getting songs on the air very seriously. On one level, we wanted them to be listenable, and we definitely did not want somebody else's record to leap out of the speakers as compared to ours*” (Jones, 2005).

Olhsson is referring to the apparent advantage elicited by exposing listeners to musical programmes[1] that are louder than the preceding programme (Packwood, 1974; Schubert, 2004a). This preference for loudness is said to be related to intrinsic human survival skill, in which Levitin (2006) proposes that the auditory system is the fastest and most important of our five senses in determining changes in our environment. However,

[1]Programme can mean a music album or single, feature film or film short, television or radio broadcast, minus the advertisements that would be considered short self-contained programmes.

while inherent loudness may influence listener preferences under certain conditions Croghan, Arehart and Kates (2012), and Maempel and Gawlik (2009) suggest that we must also account for:

> *"… aspects of aesthetic impression, communication of brand values, intelligibility, recognition, and listening convenience in different situations and at varying degrees of receptiveness".*

Loudness preference may be correlated with factors such as gender (Kellaris and Rice, 1993; Kamenetsky, Hill, and Trehub, 1997), personality (Kantor-Martynuska, 2009), and the spectral energy and variations in pitch in the musical piece (Deutsch, 1999 pp. 99-101). The multifaceted psychological aspects of loudness preference highlight the complex nature of the subject, but are ultimately beyond the scope of this research.

This thesis tests the effect of DRC on musical signals. The experimental approach incudes analysing both simple and musical signals both numerically and in listening tests with human participants. Specifically, different DRC configurations designed to minimise the effects of nonlinear signal processing are tested against listener preferences; the potential impact of musical signals subject to DRC on listener fatigue is explored; and the propensity for DRC to degrade signal quality is quantified. Based upon these findings, new DRC configuration recommendations are formulated that are intended to produce the least distortion, minimise fatigue, and maximise listener preference.

1.2 Gaps in Knowledge

i. It is unclear from existing research whether listening to music subject to heavy DRC leads to listening fatigue. Anecdotal evidence suggests that heavy DRC may cause listening fatigue and eventual hearing damage (Rumsey, 2008; Jones, 2005; Levine, 2007; Vickers, 2010; Katz, 2007). Stone et al. (2009) showed that increased DRC on simultaneous speech signals impaired attention/concentration in certain scenarios. However, no published studies exist that use musical signals as test stimuli.

ii. It is unclear from existing research which DRC configurations (DRC magnitude, application point and type) will lead to the most preferred music. The loudness of music may influence listener preference under certain conditions (Packwood, 1974; Schubert, 2004a; Croghan, Arehart and Kates, 2012; Cullari and Semanchick, 1989; Neuhoff, McBeath, and Wanzie, 1999; Barrett and Hodges, 1995; Ronan, Sazdov and Ward, 2014). However, no significant correlation links commercial music success and loudness (Vickers, 2011; Viney, 2008). There are several research projects testing the configuration of DRC and resultant perceptions (Maddams, Finn and Reiss, 2012; Giannoulis, Massberg and Reiss, 2012; De Man and Reiss, 2013; Pestana and Reiss, 2014; De Man et al., 2014; De Man et al., 2015; Ma et al., 2015). However, these studies approach DRC configurations on current and historical practice, rather than developing best practice for minimising the negative consequences of DRC. Subgrouping practice has also been investigated to determine reasons for subgrouping, along with exploration of instrument subgroup configurations (Ronan, Gunes and Reiss, 2017). However, there are currently no known published studies that examine DRC configurations specifically comparing DRC applied at specific points (track, subgroup and full-sum) for the purpose of reducing intermodulation distortions. Using speech signals, Stone et al. (2009) found that the application of DRC prior to summation produced less distortion relative to when applied after summation; therefore, one might expect that changing the configuration and degree of DRC employed for musical signals may affect distortion and therefore listener preferences.

iii. The exact impact of DRC, a nonlinear and consequently difficult-to-model process, on real music signals is not well understood. Existing studies suggest that the use of DRC for loudness maximisation compromises audio fidelity (Aarseth, 2012b; Croghan, Arehart and Kates, 2012; Essling, Koenen and Peukert, 2014; Wendl and Lee, 2014; Deruty and Tardieu, 2014; Kirchberger and Russo, 2016; Vickers, 2010; Viney, 2008;

Pestana and Reiss, 2014; Gorlow and Reiss, 2013). Stone et al. (2009) found that DRC applied to the spoken word affects the temporal contrast (amplitude variation over time), the spectral contrast (amplitude variation of the frequency spectrum) of the signal and is dependent on the number of signals having DRC applied simultaneously, onset and release speed, and DRC magnitude. Whether these observations generalise to musical signals is unclear.

iv. There are no known evidence-based studies designed to identify the optimal DRC configuration for minimising fatigue, maximising listener preference, and preserving signal quality that address DRC magnitude, point, and type parameters.

1.3 Research Questions

A. Does listening to music subjected to heavy DRC lead to listening fatigue, evidenced by inferior performance in unrelated cognitive tasks? (addressing gap in knowledge *i*).

B. Which DRC configuration leads to the most significant rate of listener preference for a given piece of music? (addressing gap in knowledge *ii*).

C. What is the quantitative impact of the nonlinear properties of DRC on real music signals? (addressing gap in knowledge *iii*).

D. Can DRC configuration manipulation minimise fatigue, maximise listener preferences, and quantitative signal quality? (addressing gap in knowledge *iv*).

1.4 Contributions to Knowledge

a. Statistical analysis of listening test results demonstrate that heavy DRC magnitudes may result in increased listening fatigue, though the results are inconclusive. These findings are evidenced by inferior performance in a cognitive task, compared to moderate magnitudes. (addressing research question *a*).

b. Statistically significant listening test results demonstrate that the manipulation of the DRC process chain leads to the most preferable music. Specifically, that DRC is preferred when applied to fewer signals simultaneously (i.e. track DRC is preferred over

Subgroup DRC, and both are preferred over full summation DRC). (addressing research question *b*).

c. The nonlinear properties of DRC applied to real music signals are similar to those described in Stone et al. (2009) for the spoken word. Specifically, analyses demonstrate that DRC affects the temporal and spectral contrast of the compressed signal, dependent on the number of signals having DRC applied simultaneously and DRC magnitude (note: DRC onset and release speed were not tested). (addressing research question *c*).

d. DRC configuration for maximising listener preference, and preserving signal quality may be achieved by manipulating the point in the mix chain where DRC is applied, *viz*., it is preferable to apply DRC to fewer signals simultaneously, use compression rather than limiting, and to apply light to moderate DRC over heavy or no DRC. (part-addressing research question *d*).

1.5 Structure of this Thesis

Chapter 2: This chapter is a theoretic review of concepts associated with historical DRC usage and factors influencing the proliferation of the 'loudness war'.

Chapter 3: This chapter discusses the fundamental aspects of digital signal processing theory, particularly nonlinear signal processing. This chapter also the examines fundamental elements of the psychoacoustic apparatus of hearing and transduction of sound relative to loudness perception.

Chapter 4: This chapter reviews the DRC design and functionality.

Chapter 5: This chapter reviews the literature associated with psychoacoustics and signal analysis studies.

Chapter 6: This chapter contains the first experiment, designed to investigates how DRC configuration affects listener preference for a given piece of music.

Chapter 7: This chapter contains the second experiment, designed to investigate auditory fatigue.

Chapter 8: This chapter contains the third experiment, designed to investigate the effects of DRC on simple sine wave stimuli and relate the findings to further tests using musical stimuli.

Chapter 9: This chapter is a discussion and critical analysis of the combined findings of the experimental sections, including comparisons and explanations relative to the research questions. This chapter will conclude with recommendations for further studies.

Chapter 10: This chapter is a summary of the key findings and their wider implications. Closing the document is a review of the contributions to knowledge.

Chapter 2:

Uses and Abuses of DRC – How DRC has been used over time

2.1 Introduction

This section reviews DRC, looking at the theoretical origins of the 'loudness wars' and the issues associated with them. Finally, there will be a more in-depth look at some of the specific issues that lead to experiments on auditory fatigue and DRC preference.

Two overlapping domains in which DRC research is conducted are speech intelligibility for hearing aid design (Kates, 2005; Tan and Moore, 2004; van Buuren, Festen, and Houtgast, 1999; Stone and Moore, 2003; Neuman et al., 1998) and DRC for controlling the loudness of musical signals (Giannoulis, Massberg and Reiss, 2012), potentially at the expense of audio fidelity (Aarseth, 2012b; Croghan, Arehart and Kates, 2012; Essling et al., 2014; Wendl and Lee, 2014; Deruty and Tardieu, 2014; Kirchberger and Russo, 2016; Vickers, 2010; Viney, 2008; Pestana and Reiss, 2014; Gorlow and Reiss, 2013). The algorithms and optimal settings in each domain differ. This research focuses on DRC used for maximising the loudness of musical signals, rather than aesthetic effects. Current hearing aid research is also summarised where relevant.

2.2 Theoretical Usage of DRC

The origin of DRC use is difficult to determine precisely. Bell Labs and RCA (circa 1932) may have developed DRC in the first experiments developing stereo reproduction for the film industry, a technology thought to be of little interest for music production. Sound recorded optically onto cellulose film had limited physical space available for the audio compared to other recording media of the era, and therefore required controlled dynamic range (Chipman, 1931). If the signal level exceeded a certain threshold, manual adjustments were made to the amplitude to compensate (compression) (Fielding, 1967, p.213; Milner, 2009). An AC signal comprising several control tones was recorded simultaneously alongside the soundtrack, logging amplitude adjustments of the 'sound track'. Upon playback the recorded amplitude control tones passed through a photoelectric cell that adjusted the signal level of the playback amplifiers (expansion) (Milner, 2009; Jacobs, 2012, pp.5-34).

As the electronics of the recording apparatus improved, becoming increasingly capable of capturing the natural dynamics of ensembles and their environments, dynamic control became increasingly essential for the protection reproduction equipment unable to handle the extreme dynamics (Newell, 2000; Borwick, 2001; Izhaki, 2008; Killion, 2009). Table 2.1 compares typical dynamic ranges of storage/playback media referenced to a typical orchestral performance (Davis, 2007, p.767; Bohn, 2011). Clearly, storage/playback media alone require DRC.

Medium	***Dynamic Range***
Optical/magnetic film	50 - 70 dB
Analogue tape	70 dB
FM broadcast	60 dB
AM broadcast	50 dB
LP record	65 dB
Consumer cassette tape (noise reduction)	60 dB
Compact disc	95 dB
Typical orchestral performance	104 - 106 dB

Table 2.1: Typical dynamic range of storage/playback media compared to that of a live orchestra.

As technology advanced the inherent noise of the electronics and storage media fell, while dynamic range decreased. Portability improved concurrently (Sterne, 2006), introducing increased background noise. For example, rumbling of tyres on the road, or people talking in the background, rendered the quietest parts of the program inaudible. Motown records famously installed car speakers in their studios to optimise mixes for the poor dynamic range of automobile reproduction (Smith, 1999, p.124). Consequently, augmented levels of DRC increased as consumer's primary listening environment changed from the home to cars and portable music players (Katz, 2007). When digital recording and playback arrived, the automobile was possibly the primary listening environment, giving way to personal playback systems such as the Sony Walkman and ultimately the iPod and other

MP3 players. Of course, personal, low-quality headphones used in noisy environments called for increased levels of compression, which anecdotally sound better with increased DRC because of their limited playback dynamics (Katz, 2007).

2.2.1 Signal Control

With the increased dynamic capabilities of Compact Disc (CD), the physical bounds of the medium were no longer a barrier to increasing the dynamics of music productions. Digital production introduced peak normalisation, capable of maximising signal amplitudes without digital overload (Shelvock, 2012). The combination of peak normalisation with DRC raises the average signal level proportionally to decreased signal dynamics.

> *"The compact disk became the catalyst for the accelerated digital loudness race. Peak normalization is the fuel that keeps the motors running."* (Katz, 2007, p.169).

Digital media are capable of higher definition and larger dynamic capacity than analogue counterparts. Unfortunately, the average dynamic range of popular music productions has slowly decreased over the lifespan of recording, despite advancing technological capabilities (Sterne, 2006). Deruty and Tardieu (2014) analysed music level and level variation of 4500 commercially and/or critically successful songs between 1967 and 2011. They also analysed a large range of DRC devices using various measures: RMS, crest factor[2], European Broadcast Union (EBU) 3341 (2010) integrated loudness, EBU 3342 (2010) loudness range, and two bespoke measures: High Level Sample Density (HLSD) that measures the proportion of samples above −1 dBFS after normalisation, and Peak to RMS Regression Coefficient (PRRC) a descriptor intended to describe micro-dynamics.

[2] Crest Factor is the ratio of a signal's absolute peak amplitude to RMS, a measure of the 'peakedness' of a signal (Benson, 1988; Katz, 2007).

They found they could deduce DRC usage from music signals, with findings including that: limiters equate to 'high RMS or HLSD values'; compressors using fast attack equate to 'low crest factor or PRRC values but normal RMS or HLSD values; compressors using slow attack equate to 'low EBU 3342 (2010) only' (Deruty and Tardieu, 2014).

Deruty and Tardieu (2014) determined that the loudness war emerged between 1988-1990 and 2004-2007. They also deduced other technology trends that likely influenced the perception of a loudness war, such as the popularity of low fidelity (lo-fi) independent bands in the late 1980's. The authors conclude:

> *"While the loudness war has indeed made mainstream music louder, transients less salient, decreased 'naturalness,' macro-dynamics remain practically untouched. In other words, there are still pianissimi and fortissimi in recent mainstream music …"* (Deruty and Tardieu, 2014).

2.2.2 Signal Maximisation

Eminent audio mastering engineers who have been active since the 1960s, such as Grammy Award winners Bob Ludwig[3] and Bob Katz[4] (AllMusic.com, 2016), suggest that the music industry has historically used heavy DRC to maximise the inherent loudness of music to grab the attention of listeners, at the expense of objective audio fidelity (Katz, 2007; Aarseth, 2012; Jones, 2005; Brownstein, 2012; Croghan, Arehart, and Kates, 2012). This 'louder is better' viewpoint derives from the apparent preference of listeners for music productions that are louder than similar productions (Packwood, 1974; Schubert, 2004; Croghan, Arehart, and Kates, 2012).

---

[3] Led Zeppelin, Queen, Rush, Jimi Hendrix, The Police, Bryan Adams, Paul McCartney, Eric Clapton, the Rolling Stones, Def Leppard, Nirvana, The Who, Guns and Roses, Daft Punk and over 3,000 other credits

[4] Foghat, Bombay Dub Orchestra, David Chesky, Cathedral, Duke Ellington, Bo Diddley / Anna Moo, Ella Fitzgerald, Ian Gillan, Peggy Lee and over 500 other credits

Four separate studies, Viney (2008), Vickers (2011), Deruty and Tardieu (2014) and Oehler, Reuter, and Czedik-Eysenberg (2015) investigated potential correlation between production success and dynamic range or loudness.

Viney (2008) investigated the correlation between commercial success and loudness of 30 'Music Week' chart singles. Commercial success was based on correlations from a 19-month period (May 2007 to November 2008), between top positions, top dates, and weeks in charts for sales, radio and TV airplay. The measure of DRC was based on an R128 (LU) average loudness measurement using a DK Technologies MSD600M++ audio level meter (DK Technologies, 2007). The study found no significant correlation (no probability results made available) between commercial success and measured average loudness.

Vickers (2011) analysed music chart ranking and sales of 173 albums from the 'Billboard 200' year-end charts (2002 to 2009) against dynamic range; based on measurements from the 'Unofficial Dynamic Range Database'[5] which utilises the 'TT Dynamic Range Meter' (a quasi-crest factor measure of the loudest 20% of the peak signal to RMS) (Tischmeyer, 2009). Vickers found no significant correlation between music sales and the use of DRC in popular music.

Oehler, Reuter, and Czedik-Eysenberg (2015) analysed 1160 songs of the German 'Top-40' year-end charts, from 1965 to 2013, to verify trends in DRC usage and potential correlation to chart ranking. They measured for loudness using: ITU-R BS.1770 Loudness K-weighted Full Scale (LKFS); RMS; dynamic range using the 'TT Dynamic Range Meter'; and the ratio of high and low frequencies at a cut-off frequency of 183 Hz (approximately the upper limit of the Equivalent Rectangular Bandwidth (ERB) with a centre frequency at 161 Hz), which may be an indicator of DRC. The use of the final measure was based on the findings of Ortner (2012) (master thesis unavailable in English) who analysed evolution

[5] www.dr.loudness-war.info

of loudness across frequency bands of more than 10,000 successful popular music recordings over a 60-year period. Ortner reported a notable increase of energy with an ERB centre frequency of 121 and 161 Hz.

Oehler et al. (2012) found that the year of recording significantly correlated with the different loudness parameters: significant increase in loudness LKFS, RMS, low-frequency content, and decreased dynamic range, substantiating the general trend found in other research (Milner, 2009; Vickers, 2010, 2011; Katz, 2007; Deruty and Tardieu, 2014). They did not find a relationship between the trend and chart ranking. These findings seem to corroborate the findings of Vickers (2011) and Viney (2008).

Deruty and Tardieu (2014) and Oehler, Reuter and Czedik-Eysenberg (2015) found two important contradictions to other research on the loudness trend. Deruty and Tardieu found that between 1984 and 1989, before the use of limiters, there was a stylistic trend toward low-fidelity music production, separate from the loudness trend. They also found that the loudness trend seemed to stabilise and perhaps reverse slightly during the period 2004–2011. However, Oehler, Reuter and Czedik-Eysenberg dispute that the trend peaked in 2004, showing the trend of increased loudness and decreased dynamic range to at least re-establish from 2011 to 2013.

Experimental evidence in support of the ‘louder is better’ argument (Giannoulis, Massberg and Reiss, 2012) for audio indicating that louder oration is more compelling (Ronan, Sazdov and Ward, 2014), and that a relationship between loud music and elicited emotion exists (Schubert, 2004b; Vickers, 2010).

Schubert (2004b) devised a computer-based experiment that tracked the perceived emotion conveyed by four romantic musical pieces using 67 volunteers that spanned a wide demographic range of age, gender, and musical expertise. The study found a significant positive correlation between loudness and both valence and arousal response dimensions, although it is also likely that loudness must be modulated to induce corresponding changes in emotional affect.

Cullari and Semanchick (1989) tested whether the loudness of music influences listener preferences. A significant positive correlation between preferred volume level and music rating was found, i.e. the more subjects enjoyed a piece of music, the louder they preferred listening to it.

Hoover and Cullari (1992) expanded on Cullari and Semanchick's (1989) experiment, to investigate the differences in loudness perception between 9 musicians and 16 non-musicians. Results showed musicians were less accurate than non-musicians at matching loudness for six of the ten excerpts, though the musician's responses exhibited lower dispersion. However, both musicians and non-musicians showed greater loudness matching accuracy with the genre that they were most familiar with. The apparent inconsistency between the findings of the 1989 and 1992 experiments were not discussed in detail by Hoover and Cullari (1992).

Considering the psychoacoustic principles that underpin loudness preference, Neuhoff, McBeath and Wanzie (1999) suggested that in a natural listening environment a sound source that increases in loudness is normally getting closer, potentially alerting the listener to a threat (or an opportunity), whilst a sound source getting quieter signals that it is moving away and therefore less important. However, Neuhoff, McBeath and Wanzie (1999) considered sounds that dynamically change in loudness, rather than the stable loudness inherent in heavily compressed signals. This raises the possibility that the so-called 'loudness advantage' gained from the use of DRC by sound engineers is only effective up to a point (perhaps during the transition from one production to another), and in fact that music that varies in dynamic loudness may be more attractive to listeners (Croghan, Arehart and Kates, 2012; Taylor and Martens, 2014).

Barrett and Hodges (1995) investigated if there was a significant difference in music loudness preference amongst middle school and college students. Participants in the study included 40 middle school (11 to 13 years old) students (MS) participating in musical studies, 40 college music students (CM), and 40 college non-music students (CN). Each group had equal numbers of male and female subjects. Subjects listened to CD-quality music from six genres (country, classical, chant, rap, jazz, and heavy metal) over headphones. The study found significant loudness preferences related to genre with heavy metal and jazz preferred loudest, while there was no significant difference between the other genres: heavy metal (84.5-96.6 dB); jazz (77.2-88.8 dB); country (65.7-74.5 dB); classical (65.7-78.0 dB); chant (69.2-79.9 dB); and rap (69.3-82.9 dB).

The study also found that there were significant differences in loudness preference between groups. The college non-music students preferred lower playback levels (except chant) and there was no significant difference between college music students and middle school students who preferred the loudest levels, perhaps showing a relationship with musical study and loudness preference. The study also found that there was a significant difference between the preferred listening levels of male and female subjects, with middle school males preferring lower levels, but fewer differences between college male and female students. Interestingly, male music students in middle school and college always listened to music louder than their female counterparts, and college non-music males always preferred to listen at lower levels than college non-music females.

Wendl and Lee (2014) examined for correlation between crest factor and perceived quality with application of DRC (limiting only) to three music genres (rock, electronic and jazz). The two-part test used audio engineering students with normal hearing (self-reported). The first part tested 15 subjects for loudness perception. The second part tested 15 subjects for quality perception; some subjects participated in both tests. Results demonstrate correlation between best perceived quality and the largest crest factor (Fig. 2.1). However, correlation in the rock genre was linear, whereas linearity deviated slightly between 12 and 14 dBFS for jazz and the electronic genre results were inconclusive. Importantly, Wendl and Lee (2014) also compared their crest factor results from test one to the EBU R128 (2011) programme loudness measurement Loudness Units Full Scale (LUFS) and visually found the two measurements to have a linear correlation.

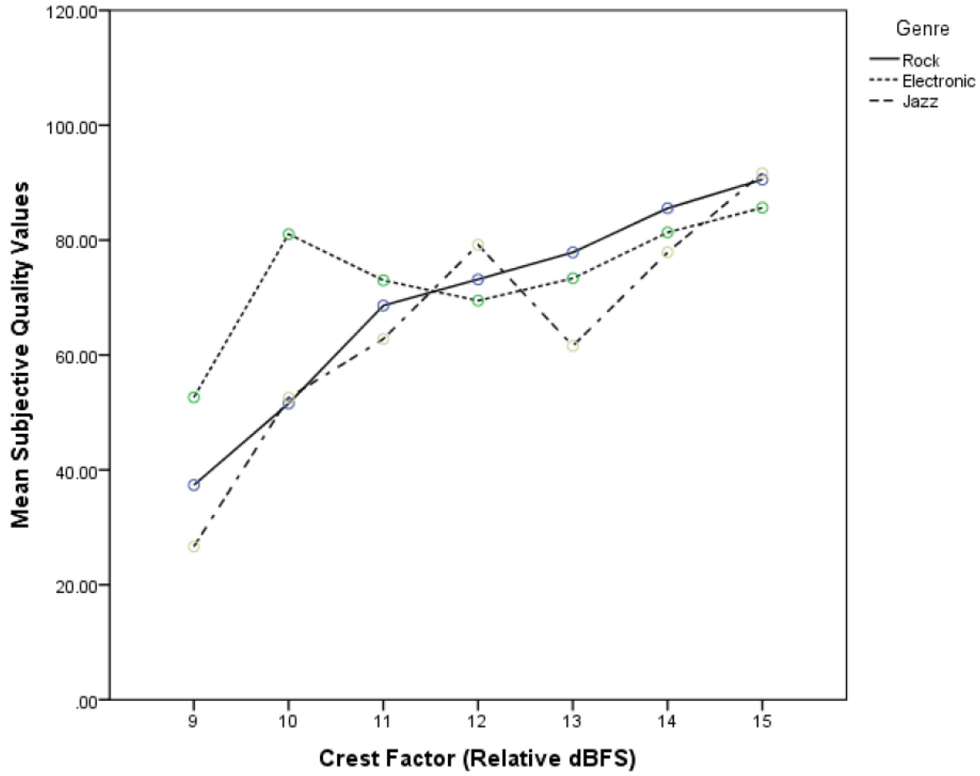


Figure 2.1: The combined mean results from Wendl and Lee (2014) showing subjective quality vs. crest factor for each music genre.

It is worth noting that Wendl and Lee use a 100-point quality scale, known to impart ‘response mapping bias’. Zielinski, Rumsey, Bech (2008) advise: "*The magnitude of the observed bias ranged up to approximately 20 points on the 100-point scale, which corresponds to a whole category change in terms of the quality labels used along the scale*." Zielinski, Rumsey, and Bech, also stipulate that the 100-point scale also introduces ‘contraction bias’, a conservative tendency by subjects in using the grading scale because: “… *listeners normally avoid the extremes of the scale* …”. We also ask the question, on what basis did participants judge quality? Reiss (2016) recommends training listeners so they can distinguish resolution; perhaps this is also true for subjective quality judgements? An anonymous mixing engineer provided experimental stimuli, prior to mastering. This may introduce bias regarding the application of DRC during the mixing process. Did the mix engineer apply DRC prior to application of DRC for the testing process and if so, was the level of DRC applied consistently for all tracks and instruments? Was DRC applied to subgroups or the master-buss? Each scenario could cause the additional DRC to perform unpredictably.

2.2.3 Artistic use of DRC

Typical DRC usage in music production renders musical signals at a more consistent level, levelling out the dynamics. Lowering the higher output portions of a performance can also aid in overcoming a high noise floor, particularly important in analogue systems. Controlled peaks allow application of additional gain, therefore maintaining signals within ideal working ranges and avoiding overload distortion (Case, 2007).

Another benefit of reducing the dynamics of an instrument is that the overall signal level and the perceived loudness can rise (Case, 2007; Izhaki, 2010). This gain in perceived loudness can transform how the audience perceives a performance. For instance, a compressed snare drum will tend to sound as if hit harder than an uncompressed snare drum because a properly set device sharpens yet controls the transient and controls the decay. Shaping the envelope in this way essentially allows raising the signal level and preserving clarity (Mynett, 2017). Configured slightly differently, the same DRC can soften the attack and artificially lengthen the decay, giving it a type of artificial tone. The opposite holds true as well; compression can make programmed music sound more natural (Izhaki, 2010; Case, 2007).

Engineer/producer Joe Meek[6] (AllMusic.com, 2016), was a pioneer in the use of DRC as an artistic tool in the 1950's and 1960's (AllMusic.com, 2016; Milner, 2009). Rather than simply using DRC to dampen the unwanted amplitude peaks of a signal, Meek discovered that through application of heavy DRC, *"he could create effects that seemed like they were jumping out of the speakers"* (Milner, 2009, p.154). This effect alters the natural

[6] The Tornados, The Honeycombs, The Shadows, The Ventures, Gene Pitney

dynamic envelope of a signal, or in other words the amplitude variations that happen within individual notes (or hits), the micro-dynamics. Altering micro-dynamics alters an instrument's timbre (Izhaki, 2010).

As DRCs are nonlinear devices, altering the envelope of an instrument will introduce harmonic and intermodulation distortion that can help an instrument stand out in a mix (Case, 2007; Izhaki, 2010). Not only can one apply this to individual instruments such as electric guitar or the bass drum, it can also provide great effect when applied to macro-dynamic aspects of mix such as choruses or solos (Case, 2007; Mynett, 2017).

An effect known as pumping, associated with an unnaturally long release time of a compressor can produce an interesting effect. During the release stage of DRC, a release time set slightly longer than the natural envelope of the instrument can artificially amplify other instruments. For instance, a snare drum with a slightly long release time may cause hi-hat or cymbal to pump (artificially rise and fall in the mix) (Case, 2007; Izhaki, 2010).Electronic Dance Music (EDM) producers sometimes intentionally use this for effect. 'Breathing' is an artefact associated with gates and expanders when the noise floor is close to the threshold (Case, 2007; Izhaki, 2010).

2.3 Additional Factors Driving the Loudness War

This section other discusses factors leading to and driving the progression of the loudness war.

2.3.1 Analogue and Digital Metering Practice

Analogue recording engineers used Volume Unit (VU) meters to monitor recording/mixing levels. The VU meter was useful for analysing level due to circuitry designed for a dynamic time response of close to 300 ms and an over-swing of not more than 1.5%, thus giving a rough indication of average signal level and limited peak signal level. This meant that the average signal level was, for the most part, consistent across publications due to standardised ballistics (Katz, 2007; Nisbett, 1974).

In the mid-1970s, Dr. Thomas Stockham of Soundstream Inc. made the first 16-bit digital recording. Not long after, recording digitally onto tape overtook analogue in professional recording studios (Schoenherr, 2004). Unfortunately, analogue metering practices were applied in their entirety to digital.

Digital Peak Programme Meters (PPM) do not measure the loudness of a signal, but are a representation of the peak sampling amplitude. VU meters on the other hand do not represent digital signals well because their mechanical inertia means they are less likely to register overloads; analogue magnetic tape was fairly forgiving (sonically) of such overloads whereas digital is not (Newell, 2000; Katz, 2007; Borwick, 1996).

In the early days of the CD, mastering engineers would dub analogue tape at 0 VU set to -20 dBFS. This would allow plenty headroom for the natural crest factor. Therefore, the intrinsic loudness of early CDs was generally consistent. However, the inventors of the digital production systems abandoned the VU meter for the PPM. The intrinsic differences in functionality meant loss of standardised average levels (Katz, 2007; Nisbet, 1974).

Discontent for loudness extremes became evident in the popular music industry. For instance, fans of the music group Metallica published an online petition in 2008, demanding that the record label '*re-mix or re-master Death Magnetic*' (2008). The complaint was based on fan observations that the version released for the video game 'Guitar Hero' (Harmonix Music Systems, 2005), that used the music for the soundtrack, seemed to have higher fidelity than the CD version (Shepherd, 2008; Michaels, 2008). The mastering engineer for Death Magnetic, Ted Jensen, said that the CD sounds different from the 'Guitar Hero' version due a technique called 'brick wall limiting' (referred to as limiting) used for mastering the final production mix (Metallicabb.com, 2008).

Table 2.2 below compares the peak amplitude and RMS amplitude of Metallica's '*The Day That Never Comes*', (Metallica, 2008) the track with the highest amplitude on the Death Magnetic CD. Figure 2.2 shows the dynamic variations of the PlayStation 3 Guitar Hero version, and Fig. 2.3 shows the lack of dynamic variation on the CD version.

PlayStation 3 Guitar Hero version	*Channel 1*	*Channel 2*
Peak Amplitude	-1.75 dBFS	-1.65 dBFS
Peak RMS Amplitude (AES std.)	-6.97 dBFS	-6.33 dBFS
CD version	***Channel 1***	***Channel 2***
Peak Amplitude	0.00 dBFS	0.00 dBFS
Peak RMS Amplitude (AES std.)	-1.11 dBFS	-1.07 dBFS

Table 2.2: 'The Day That Never Comes' Peak and RMS analysis.

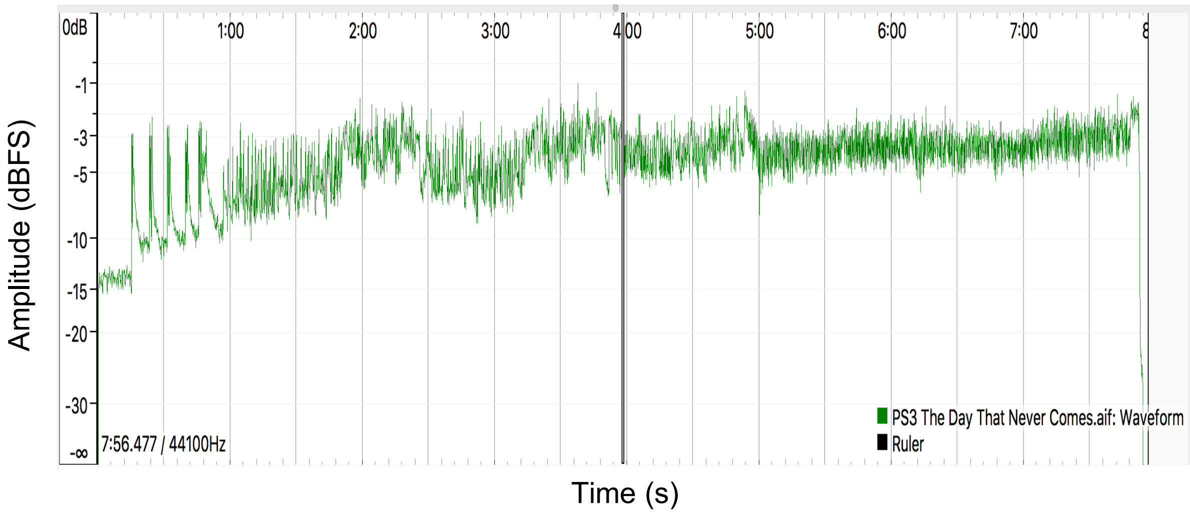


Figure 2.2: PS3 Guitar Hero version of 'The Day That Never Comes'; showing the peak amplitude (dBFS) of entire track vs. time in seconds.

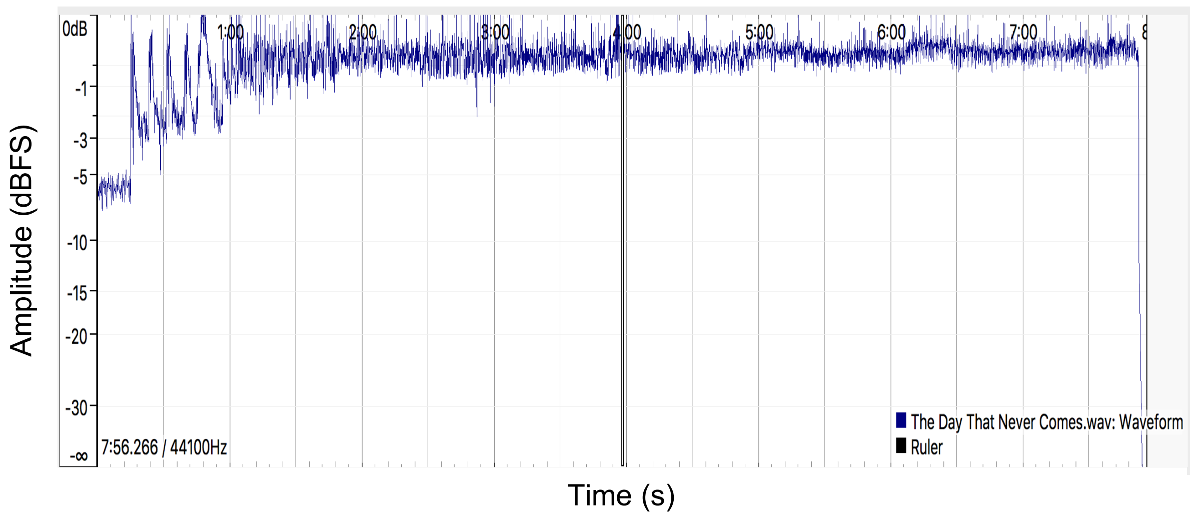


Figure 2.3: Waveform for CD version of 'The Day That Never Comes'; showing the peak amplitude (dBFS) of entire track vs. time in seconds.

2.3.2 Metering standardisation

The combination of metering ambiguity, the relatively fast transition from analogue to digital production, advanced usability/capabilities of digital, and the competitive nature of the music industry (and other media for that matter) culminated in an eventual backlash from industry and the public. Between 2008 and 2010, the American Federal Communication Commission (FCC) received 819 complaints and 4,582 inquiries from consumers regarding loud advertisements (McGrath, 2011). Differences in program loudness, particularly between television programs and commercial content in the United States and listener/viewer dissatisfaction ultimately culminated with then-President Barack Obama signing the 2010 Commercial Advertisement Loudness Mitigation Act (CALM). It required the FCC to create a regulation limiting the audio 'volume' for commercials broadcast on television (Cabot and Dennis 2011).

Before 2006 there was no international standard to measure loudness. The transition from analogue to digital television (DTV) prompted the use of loudness measurement technologies and the creation of BS.1770. In October 2010, the ITU committee maintaining BS.1770 accepted changes submitted by the EBU (EBU R128), resulting in improved calculation of loudness. This change attempted to prevent advertisers from significantly increasing the loudness of a small portion of the program by drastically reducing the loudness in other parts of the program (Cabot and Dennis, 2011). The EBU PLOUD committee revisited BS.1770 resulting in the 2011 revision, maintaining the same filtering and power management method of the original standard, but changing the way measurements were averaged and presented. The aim of the PLOUD committee was one loudness standard for the whole audio chain from production to reproduction through EBU R128, establishing one loudness level of -23 LU ±1 LU (Camerer, 2010).

ITU-R BS.1770 is an objective measurement of perceptual loudness, aligning programs to the perceptual loudness rather than program peaks (Camerer, 2010). Advanced Television Systems Committee (ATSC) RP A/85, ITU-R BS.1864 and EBU R128 are recommendations based upon ITU-R BS 1770, but are subordinate. These three recommendations are all different ‘flavours’ of the same thing, all shifting from the traditional paradigm of peak normalisation to loudness normalisation (Camerer, 2010).

2.4 Conclusion

As technology advanced from early DRC usage, the inherent noise of electronics and storage media fell, becoming increasingly capable of capturing the realistic dynamics of music performances. Augmented levels of DRC increased as consumer’s primary listening environment changed from the home to cars and portable music players (Sterne, 2006; Smith, 1999; Katz, 2007). However, Grammy award winners Ludwig and Katz express that the real factor underlying the loudness trend is that the music industry has historically used heavy DRC to maximise loudness of music to grab the attention of listeners, at the expense of objective audio fidelity. This leads us to question, is the loudness trend perpetuated by the music industry or is it a consequence of supply and demand?

Viney (2008), Vickers (2011), Deruty and Tardieu (2014) and Oehler, Reuter, and Czedik-Eysenberg (2015) all verify the trend in loudness for popular music. Deruty and Tardieu and Oehler, Reuter, and Czedik-Eysenberg found two important contradictions about the loudness war conspiracy theory. Between 1984 and 1989, before the use of ‘brick wall’ limiters, there was a stylistic trend toward low-fidelity music production, separate from the loudness trend. Also, the loudness trend seemed to stabilise and perhaps reverse slightly during the period 2004–2011. Oehler, Reuter, and Czedik-Eysenberg dispute that the trend peaked in 2004, showing the trend re-establish from 2011 to 2013. However, no significant correlation links commercial music success and loudness (Vickers, 2011; Viney, 2008; Deruty and Tardieu, 2014 and Oehler, Reuter, and Czedik-Eysenberg, 2015). So why has the trend perpetuated?

The loudness of music may influence listener preference under certain conditions. Cullari and Semanchick (1989) found that the more subjects enjoyed a piece of music, the louder they preferred it. Barrett and Hodges (1995) indicated participant preference for loudness related to genre, with heavy metal and jazz preferred loudest. Other experimental results indicate that louder oration is more compelling (Ronan, Sazdov and Ward, 2014). However, Croghan, Arehart and Kates (2012) found that heavy DRC decreased the likelihood of stimuli selection as preferred in both Loudness Un-equalised (UNEQ) and Loudness Equalised (LEQ) conditions, and yields reduction in pleasantness ratings. Croghan, Arehart and Kates examined the impact of DRC on perceived audio quality, along with loudness. Two uncompressed 13-s recordings from rock and classical genres served as source stimuli. DRC was applied to the final (summed) signal for each recording, yielding six compression thresholds (uncompressed, –8, –12, –16, –20, and –24 dBFS). Two versions of each stimulus were created, one in which loudness was not equalised between stimuli (UNEQ), and another in which LEQed. Twenty-three participants rated stimuli played in randomised pairs, within genre, on the metrics of preference, loudness, pleasantness, and dynamic range. Six hours of testing, spread over multiple visits, were completed. In both the UNEQ and LEQ conditions, the effect of compression was found to have a significant effect, for both rock and classical music stimuli, on loudness, dynamic range, pleasantness, and preference ratings (all $p \leq .01$). More specifically, these results indicated a genre-independent preference for light DRC over no DRC in the UNEQ condition. The dynamic range ratings for stimuli subject to DRC were markedly lower, showing that the principal outcome of DRC was noticeable. Heavy DRC was also found to decrease the likelihood of these stimuli being selected as preferred in both UNEQ and LEQ conditions, and to yield a commensurate reduction in pleasantness ratings.

Kirk (1956) found untrained listeners preferring lower-fidelity audio likely due to their prior listening experiences and listening preferences, which is malleable (through training) either towards objectively higher fidelity, or objectively lower fidelity recordings. Olive (2011) and Olive (2012) support Kirk’s findings, suggesting that listening preferences alter with habituation/training. They found that teenagers and college students can discern and appreciate differences in sound quality under controlled listening conditions. Reiss (2016) conducted a systematic review of 18 studies published after 1980 that evaluated listeners’ ability to discriminate high *vs*. standard resolution audio (defined in terms of bit depth and sample rate). It was reported that the ability of untrained listeners to discriminate between resolutions was surprisingly poor, but did improve significantly after training, supporting the notion that an ability to discriminate between audio quality settings can be learned. These findings support the notion that ability to discriminate audio quality standards is feasible despite previous experience.

Anecdotal evidence suggests that heavy DRC may cause listening fatigue and eventual hearing damage. Several studies demonstrate fatigue relative to noise: (Sarampalis et al., 2009) experiments indicate that background noise can have a negative effect on listening tasks, secondary simultaneous task performance, and cognitive activities. Brown's (1997) reported disruption to the temporal production tasks while performing a non-temporal task because fewer resources were available for temporal processing. Rabbit (1968) found that hearing poorer quality audio during a task can negatively impact verbal memory. However, Cassidy and MacDonald (2010) found that when subjects listened to self-selected music, there was a significant benefit to performance in both speed of play and accuracy of play. When the experimenter selected test music, there was a negative effect on speed and accuracy of play, in agreement with Stone et al. (2009), performance recovered with participants' familiarity with the distractor. Also, Cassidy and MacDonald (2010) found participants asked to estimate the time elapsed while playing the game, while listening to self-selected music had the longest time estimates, and listening to music selected by the experimenter resulted in the shortest time estimates. Droit-Volet et al. (2004) propose that temporal overestimations may indicate emotional arousal (happy, sad, angry). Danckert and Allman (2005) submit that underestimations may indicate boredom.

It is unclear from existing research whether listening to music subjected to heavy DRC leads to listening fatigue. Stone et al. (2009) showed that increased DRC on simultaneous speech signals impaired attention/concentration in certain scenarios. However, no published studies are known that use musical signals as test stimuli.

Studies suggest that the use of DRC for loudness maximisation compromises audio fidelity. Stone et al. (2009) found that DRC applied to the spoken word affects the temporal contrast and the spectral contrast. The degree of compromise is dependent on the number of signals having DRC applied simultaneously, onset and release speed of DRC, and DRC magnitude. Whether these observations generalise to musical signals is unclear. The exact impact of DRCs nonlinear behaviour is a process that is difficult to model using real music signals, and therefore not well understood. Stone *et. al.* also found that the application of DRC to speech signals prior to summation produced less distortion relative to when applied after summation; therefore, one might expect that changing the configuration and degree of DRC for musical signals would influence the level of distortion and therefore listener preferences. Similarly, Toulson, Paterson, and Campbell (2014) found that applying equivalent amounts of DRC to simple *vs*. summed sine waves affected the severity and type of distortion introduced, as reported by Le Brun (1979). Specifically, DRC applied to signals prior to summation reduces intermodulation distortion. There are several research projects testing the configuration of DRC and resultant perceptions. Ronan et al., (2015) and Ronan, Gunes and Reiss (2017) investigated subgrouping practice to determine reasons for subgrouping and various instrument subgroup configurations. Ten award-winning mix engineers expressed preference for DRC application to drum and vocal subgroups to maintain good gain structure, and organisation purposes. They also reported subgrouping decisions depended upon musical genre. Ronan et al., (2015), and Pestana and Reiss (2014) reported subgroup or full-sum DRC as preferred by sound engineers/audio mixing students. However, current studies approach DRC configurations on contemporary practice, likely built on historical tradition from the tape era, rather than developing best practice for minimising the effect of DRC in the digital era. One of the problems associated with the loudness war was the transition from analogue to digital; analogue practices were transferred to digital in their entirety (Newell, 2000; Katz, 2007; Borwick, 1996). There are currently no known published studies that examine DRC configurations specifically comparing DRC applied at specific points (track, subgroup and full-sum) for the purpose of reducing intermodulation distortions, leading to the most preferred music.

The next chapter addresses unanswered questions from the literature using three experiments: the first experiment tests whether listening to music subjected to heavy DRC leads to listening fatigue; the second experiment tests which DRC configurations (DRC magnitude, application point, and type) are most preferred; and the third experiment examines and quantifies the exact impact of DRC on real music signals.

Chapter 3:

Essential Signal Processing and Auditory Science Concepts

3.1 Introduction

This chapter contains relevant digital signal processing (DSP) theory, and explanations of some analogue signal processing techniques associated with DRC, including the basic theory underpinning compressors/limiters. There will also be an introduction to the auditory science concepts related to loudness perception.

3.2 Loudness and Signal Level Metering

Loudness is an ambiguous term used to describe the magnitude of an auditory sensation (Fletcher and Munson, 1933). Loudness is primarily dependent on the physical magnitude (sound pressure) of the sound. However, loudness is subjective and therefore it is not easy to characterise the relationship between the physical quantity or quality related to subjective impression of loudness (Benson, 1988). Loudness perception is also influenced by the following factors: frequency content; signal level; timbre; intensity; signal duration; direction of source; relative distance; temporal duration; transfer characteristics of transduction equipment; and physical characteristics of the listener (Fletcher and Munson, 1933; Benson, 1988; Stevens, 1957; Moore, 2013).

Complex models of how loudness correlates to Sound Pressure Level (SPL) and frequency were developed by prominent scientists. Fletcher and Munson (1933) developed the original equal-loudness contours by asking participants to match the loudness of discrete test tones ranging between 20 Hz and 15 kHz to a 1000 Hz test tone over a series 10 dB increments ranging between -10 dB and 130 dB SPL. The resulting was a series of phon (a unit of loudness level) contours describing how the average listener perceived loudness over the frequency spectrum. For example: if the reference tone of 1000 Hz is fixed at 40 dB SPL, a 40 phon equal-loudness contour across the frequency spectrum would be produced (Fletcher and Munson, 1933; Robinson and Dadson, 1956; Hellman, 1985; Glasberg and Moore, 2002). The shapes of the equal loudness contours vary markedly across the different studies due to variations in the testing methods, culminating in International Organization for Standardization (ISO) equal loudness contours (ISO 226: 2003) derived from the findings of Takeshima et al. (1994) (Fig. 3.1). The current ISO equal-loudness contours derive from extensive measurements from several laboratories (Moore, 2013).

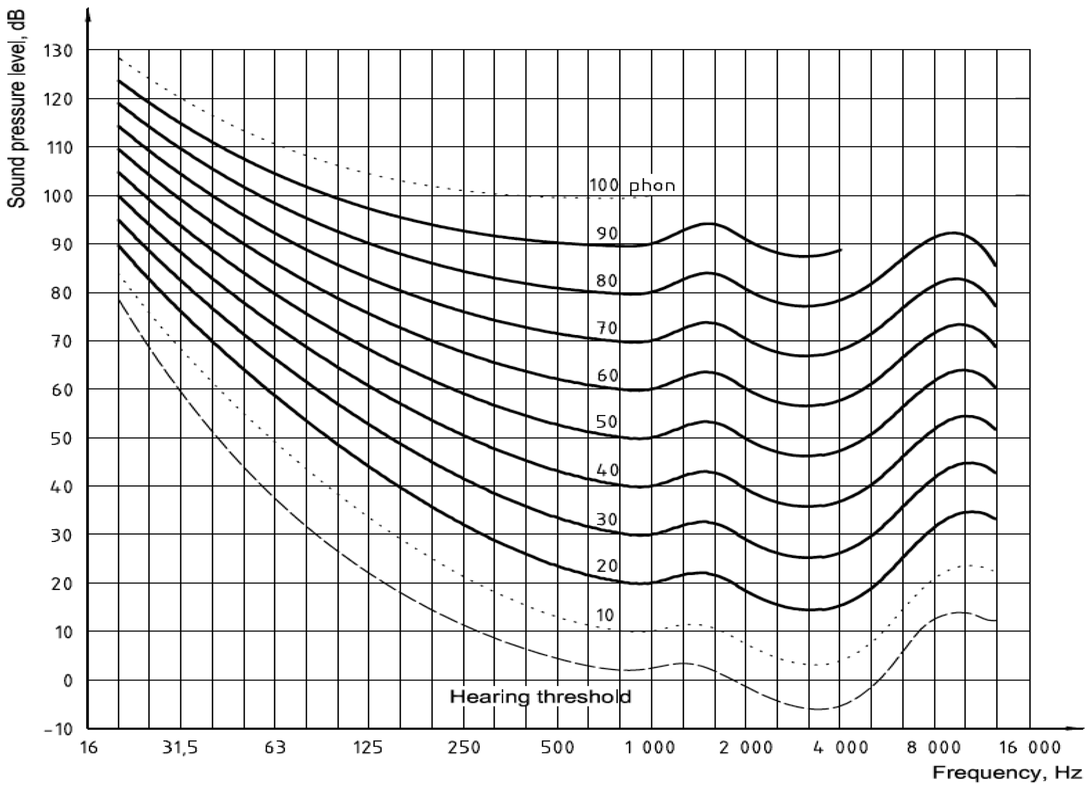


Figure 3.1: Equal loudness contours directly from ISO 226:2003(E)

For a given sound to be perceived to be twice as loud as another, the intensity must be increased by a factor of 10. However, this rule does not apply if the two sounds are not in the same critical band. Critical bandwidth has been revised over many years and now referred to as 'auditory filters', which can be represented by the Equivalent Rectangular Bandwidth (ERB) (Eq. 11) (Moore, 2013). ERB (Fig. 3.2) is described as when two sounds of equal loudness of relatively similar pitches, are sounding together, the combined signals will sound only slightly louder than the individual signals separately. The further the separation of the two signals in pitch, the louder they will sound as a combined pitch, compared to being sounded individually as two distinct pitches. The explanation for this phenomenon is the basilar membrane behaving as if it contains a bank of bandpass filters, with overlapping passbands, as suggested first by Helmholtz (1863) and then Fletcher

(1940). Each point on the basilar membrane responds to a limited range of frequencies, corresponding to a filter with a different centre frequency. Glasberg and Moore, (1990) devised a mathematical descriptor of the ERB_N as a function of the centre frequency F:

$$ERB_N = 24.7(4.37F + 1) \qquad \textbf{Eq. 11}$$

Where ERB_N is specified in Hz

F is specified in kHz

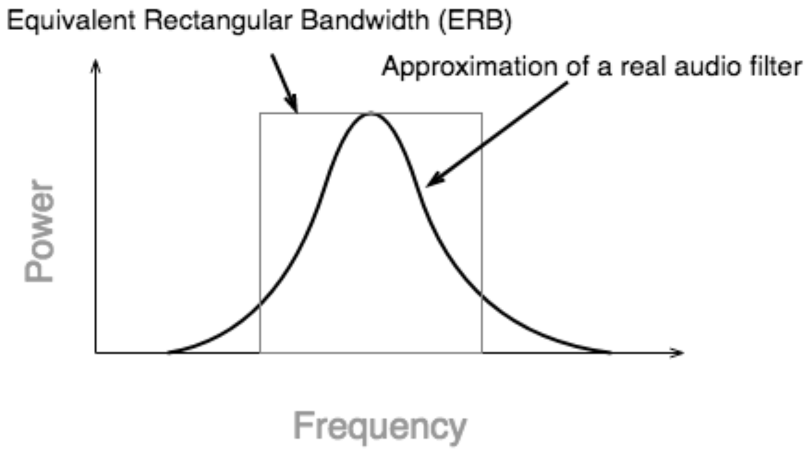


Figure 3.2: The relationship between ERB and approximated real audio filters (after Jurado and Robledano, 2007)

As an example, the ERB_N of a 1 kHz signal would be 130 Hz, equating to roughly 65 Hz either side of 1 kHz, meaning the ERB_N would be 965–1065 Hz. However, please note that this is only a rough measure, as it does not account for the magnitude of a signal. For further information on ERB_N see Moore (2013).

The sone is a unit of measurement intended to provide a linear scale of loudness based on the observation described above that a 10-phon increase in a sound level is often perceived as a doubling of loudness (Eq. 12). Stevens (1957, 1972) suggested that perceived loudness, L, is a power function of physical intensity, I:

$L = kI^{0.3}$ **Eq. 12**

Where k is a constant, dependent on the subject and the units used.

Simply put, the loudness of a given sound is proportional to the intensity raised to the power of 0.3 (Moore, 2013). Additionally, the term is expressed using amplitude or sound pressure by taking the RMS pressure raised to the power of 0.6 (Moore, 2013). Table 3.1 shows the relationship between musical dynamics, loudness and sound intensity. Note that dynamic levels in musical notation are relative rather than specific and understood as ranges or areas (Moylan, 2007). Therefore, relationships are based on ***fff*** (fortississimo) being the loudest (live classical) and ***ppp*** (pianississimo) being the quietest dynamic levels in common use. Further, the relationship does not hold for signals lower than 1 sone (Fig. 3.3).

Dynamics	***Phons***	***Sones***	***Intensity Wm^{-2}***	***Ratio I/I_0***	***Level (dB)***
fff	100	64	10^{-2}	10^{10}	100
ff	90	32	10^{-3}	10^{9}	90
f	80	16	10^{-4}	10^{8}	80
mf-mp	70	8	10^{-5}	10^{7}	70
p	60	4	10^{-6}	10^{6}	60
pp	50	2	10^{-7}	10^{5}	50
ppp	40	1	10^{-8}	10^{4}	40

Table 3.1: The relationships between musical dynamics, loudness and intensity (adapted from Pierce, 1983, p.125).

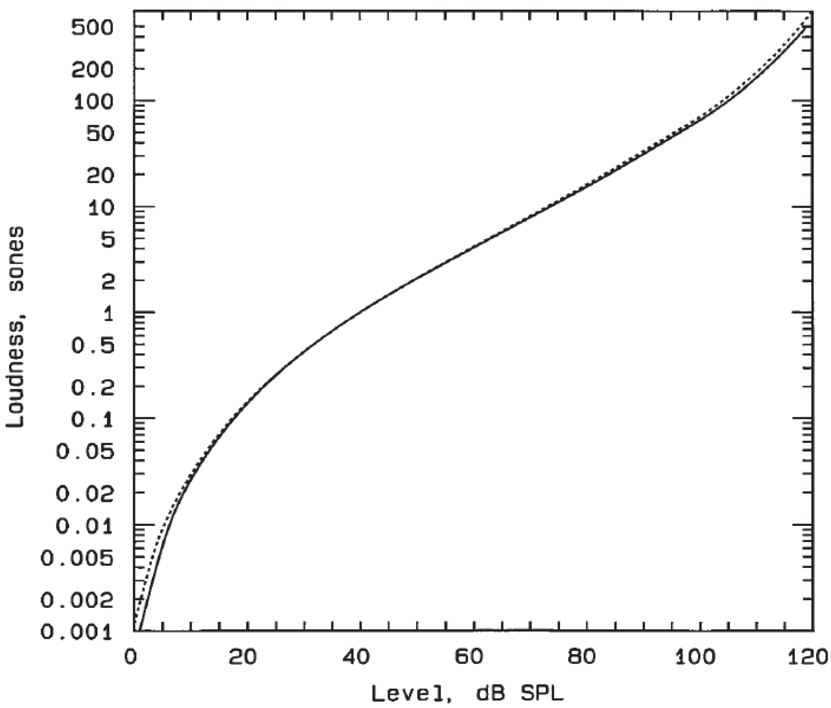


Figure 3.3: A graphical representation of the relationship between loudness in sones and SPL for a 1000 Hz sinusoid (directly from Glasberg and Moore, 2002).

Unfortunately, mechanisms governing the perception of loudness are incompletely understood, but likely arise from the interaction of many factors. Generally accepted theory suggests that loudness perception probably relates to the total neural activity evoked by a sound (Moore, 2013).

For a given neuron each neural impulse (firing) has a fixed size, the number and timing of the impulses relay information about the auditory source. In the absence of sound, the neurons fire at Spontaneous Impulse Rates (SIR) ranging from 0 – 150 per second. Neurons respond better to some frequencies than others, selectivity differs between neurons. Below around 5 kHz, neurons demonstrate ‘phase locking’; meaning the SIRs occur at a particular phase of the stimulating waveform, with temporal regularity sympathetic to the waveform. Phase locking provides inter-aural time difference information to both ears and the central auditory system, vital for localisation of sounds and our perception of auditory space (Bear et al., 2007). Neurons have one of three SIRs: 61% of neurons exhibit high SIRs of 18-250 impulses per second; 23% exhibit medium SIRs of 0.5-18 impulses per second; 16% exhibit low SIRs below 0.5 impulses per second.

SIRs correlate to the position and size of inner hair cell (IHC) synapses. The IHC stereocilia are the sensory receptors of the cochlea and transduction of BM motion into neural synapses is primarily through the IHCs (Moore, 2013). High SIR are associated with large IHC synapses, particularly those closest to the Outer Hair Cells (OHC). The OHCs theoretically influence the mechanics of the cochlea and IHCs through efferent nerve fibres (Fuchs, 2010). Likewise, low SIRs correlate to small IHC synapsis, furthest from the OHCs.

The neural threshold indicates the lowest sound level that induces increased SIR. The neural threshold for a given frequency is a given fibre characteristic frequency, which increases for frequencies either side of the characteristic frequency. Neurons have a saturation point where SIR reaches the maximum with growing intensity. High SIRs tend to occur at low thresholds and for small dynamic ranges. The most sensitive neuron may have a threshold near 0 dB SPL, while the least sensitive threshold near 80 dB SPL or more (Moore, 2013).

SIRs are relatively high when first activated by a sound and tend to decrease over time, an effect called adaptation. When the sound ceases, the SIR may drop below the normal rate. Fatigue is evident if the absolute threshold of the neuron increases after the sound has abated (Moore, 2013). Temporary Threshold Shift (TTS) resulting in auditory sensitivity after exposure to the loud sound, indicates an eventual return to the original threshold. TTS recovery time can vary from seconds to days.

Loud sounds can also induce Permanent Threshold Shift (PTS), or permanent hearing loss. Factors that influence TTS and PTS are the sound intensity level and duration of the fatiguing stimulus, frequency content, and time from cessation and threshold evaluation, i.e. the recovery interval (Moore, 2012; Plack and Moore, 2010). Post-stimulatory auditory fatigue might be fatigue of the cochlear hair cells, but TTS understanding is limited (Epstein, 2013; Plack and Moore, 2010; Scharf, 2001; Yoshida et al., 2006; Moore 2012).

Loudness perception may depend on the summation frequencies in critical bands or ERB_N, discussed in section 3.2. Even when SIR saturation from high SPLs occurs at the ERB_N centre, intensity changes are detectable from the spread of the excitation pattern by SIRs at the edges of the ERB_N and by SIR changes at the high and low ends of the ERB_N pattern (Moore, 2013).

A better understanding of loudness perception might come from understanding where SIRs happen relative to temporal intervals of phase locking during depolarisation, the positive phase of a low-frequency sound wave (Moore, 2013; Bear, Connors, and Paradiso, 2007). Phase locking occurs at approximately integer multiples of the period of a stimulating waveform. The relative levels of the components of complex signals may signal different orientations of phase locking. Phase locking is not evident for frequencies exceeding 4-5 kHz, perhaps faintly up to 10 kHz. The information contained in the neural firing rates is known to be sufficient for intensity discrimination, but the more central parts of the auditory system probably limit loudness perception (Moore, 2013). Studies of both animals and humans show the vestibulocochlear nerve, the eighth paired cranial nerve, also called Auditory Nerve (AN), transmissions are poor reproductions of auditory stimulus for frequency and amplitude. Theory suggests auditory periphery may be explicitly oriented for species-typical vocal sounds rather than replicating all sounds uniformly (Bear, Connors, and Paradiso, 2007).

The AN consists of afferent and efferent nerve fibres. The afferent fibres transmit impulses to the brain stem and Cochlear Nucleus Complex (CNC). The efferent fibres return signals from the CNC to the organ of Corti (Rees and Palmer, 2010). AN fibres terminate within the CNC and segregated by tonotopic representation (Moore et al., 2010). The segregated synapses shunt along parallel ascending tracts to the supra-temporal cortex, which represents the primary auditory cortex and the associated areas surrounding the primary auditory cortex (Rees and Palmer, 2010; Gazzaniga, Ivry, and Mangun, 2002; Bear, Connors, and Paradiso, 2007).

Descending auditory pathways from various regions of the cerebral cortex transmit signals via the CNC back to the cochlea. The descending circuitry likely provides a variety of functions, such as feedback that modifies information sent to the ascending pathways. Modifications control what information reaches consciousness such as selective attention functions (ability to focus on a specific audio source). Modifications may also improve cognitive extraction of salient information. Descending circuitry may be protective, for example, minimising damage from intense sounds (Rees and Palmer, 2010). The efferent system may attenuate signal levels to the afferent system by regulating the gain and sharpness of tuning of the IHCs and OHCs (Moore, 2013). Cochlear gain control is essential for interrelated issues of intensity encoding, reduction of tonal masking and performance within noisy environments (Rees and Palmer, 2010).

Additionally, efferent neural transmitters of the inner ear link directly to the cerebellum (Levitin, 2011) and the brain stem (Price, 2005). The cerebellum (the oldest evolutionary part of the brain) is known to be partially (with the parietal lobe, and the basal ganglia) responsible for motor control, coordination, and orientation to auditory stimulus in space (Levitin, 2011). The brain stem regulates heart rate, breathing, sleeping, and eating (Price, 2005). This hardwiring system provides automatic responses to environmental indicators of danger, such as loud noises, stimulating a fight-or-flight reaction, thereby increasing survivability (Levitin, 2011; Price, 2005).

3.2.1 Volume Indicator Meter

Volume Indicator (VI) meter ballistics respond at a fast rate, representative of the peaks and valleys of the voltage of a signal (Ballou, 2013). The VI meter measures power levels of audio signals and confined to use in test equipment for steady-state measurements.

3.2.2 VU

VU meters are a special form of a VI meter, having linear ballistics designed to respond at a slower rate, showing the RMS level of complex waveforms (Ballou, 2013; Benson, 1988). The VU is for monitoring material that varies in amplitude and frequency in broadcast and recording circuits. Their ballistics function essentially between signal average and peak values of complex waveforms and designed to have a dynamic response characteristics that approaches the response of human hearing.

VU meters are inaccurate loudness meters as they offer equal weighting across the frequency spectrum (Benson, 1988). Hearing is less sensitive to low frequencies than it is to higher frequencies as demonstrated in the equal-loudness contours (Fletcher and Munson, 1933; Robinson and Dadson, 1956; ISO 226, 2003). The ballistics (and therefore the accuracy) of VU meters differ between manufacturers, as a result of balancing manufacturing costs with product affordability, cheaper consoles have VU meters that are less accurate than the more expensive counterparts (Newell, 1999).

3.2.3 Sound-Level Meter

Sound-Level Meters (SLM) are the type of the VU meter used for acoustical measurements, adjustable in amplitude and time response. SLM utilise various frequency weighting curves for measurements associated with loudness, annoyance, and noisy environments for hearing risk assessment (Benson, 1988). The A-weighting (Fig. 2.7) is the most popular, and is based on the 30 phon equal-loudness contour. The B-weighting measurement is based on the 70 phon equal-loudness contour and is used for intermediate sound pressure levels. C-weighting is appropriate for high-SPLs when all frequencies contribute relatively equally, as shown as equal-loudness contours (Fig. 3.4) (Moore, 2013).

There are issues associated with the use of SLM. Firstly, they are most suited to measuring steady-state sounds of relatively long duration, and have poor transient response. Second, SLM do not satisfactorily sum the loudness of components outside ERB_N; the loudness of a complex sound is based on the energy contained within a narrow range of frequencies or spread over a wide range of frequencies. Finally, SLM do not provide perceived loudness readings directly as the meters calibration is for physical magnitude (dB) rather than a perceived loudness scale (Moore, 2013).

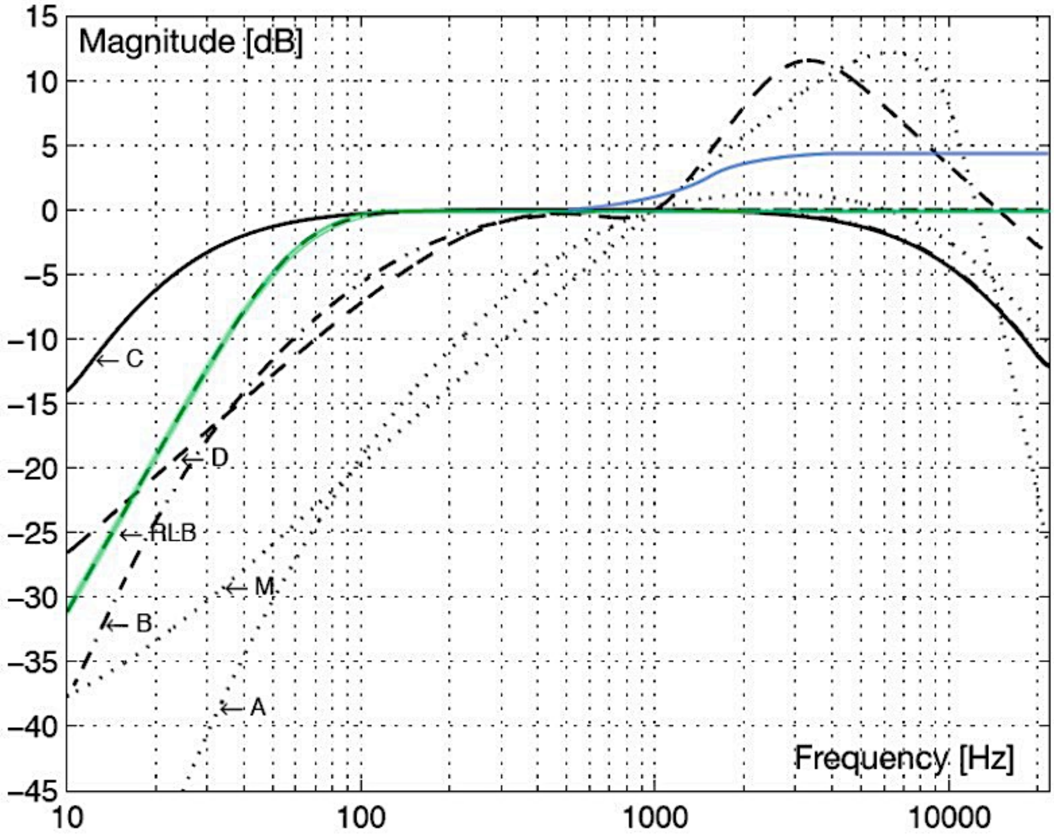


Figure 3.4: Various frequency weighting curves. K-weighting is the combination of the blue (stage 1 pre-filter) and green (stage 2 pre-filter) curves. (used with permission of Thomas Lund).

3.2.4 PPM

PPMs are historically quasi-peak program meters. The PPM is effectively a VI meter arranged to use a logarithmic scale over its working range. The PPM has rapid rise (0.1 ms to 10 ms) with slow decay characteristics (Wiley, 2003), allowing the user to easily monitor signal peaks. PPMs measure signal peaks better than VU meters.

In the digital domain, PPMs often register peak samples (the amplitude of the metered signal at the sampled interval) rather than true peak (the actual peak amplitude of the metered signal) (Fig. 2.8). Signal peak amplitudes approaching 0 dBFS can often peak between sampling instances, with higher frequency signals increasing potential error (ITU-

R BS 1770; Cabot and Dennis, 2011). Additional signal processing, particularly processing that introduces phase shifts or time offsets such as sample rate converters, filtering, or delay (Cabot and Dennis, 2011).

PPMs do not indicate loudness, unlike VU meters. The ear responds roughly to average signal levels when judging loudness, therefore RMS level is a more accurate indication of loudness to peak signal levels. RMS is unresponsive to phase shifts and therefore effectively measures the 'true energy level' of signals (Rumsey, 2008; Katz, 2007).

3.2.5 ITU BS.1770 Metering

The International Telecommunications Union (ITU) decided that the PPM meter was an inadequate measure of signals with heavy DRC applied. Although PPM design allowed for headroom (space for signal peaks without exceeding 0 dBFS), it was eventually used up by the modern loudness maximising practices (Camerer, 2010). Digital overages and therefore distortions were introduced into the signal stream. Overages are at every stage of the signal processing chain becomes very significant.

ITU-R BS 1770 metering design incorporated the Revised Low frequency, K and B-weighted filters (RBL) (Fig. 3.5) into the circuit design so the metering would be representative of general human loudness perception. The filtered metering system incorporates a high-pass filter designed to account for human insensitivity to low frequencies and the acoustic effects of the head. In multi-channel systems, the filtering is applied to all channels except the low frequency effects channel to accommodate the fact that sounds arriving from behind the listener may be perceived as louder than those from the front (Soulodre, 2004).

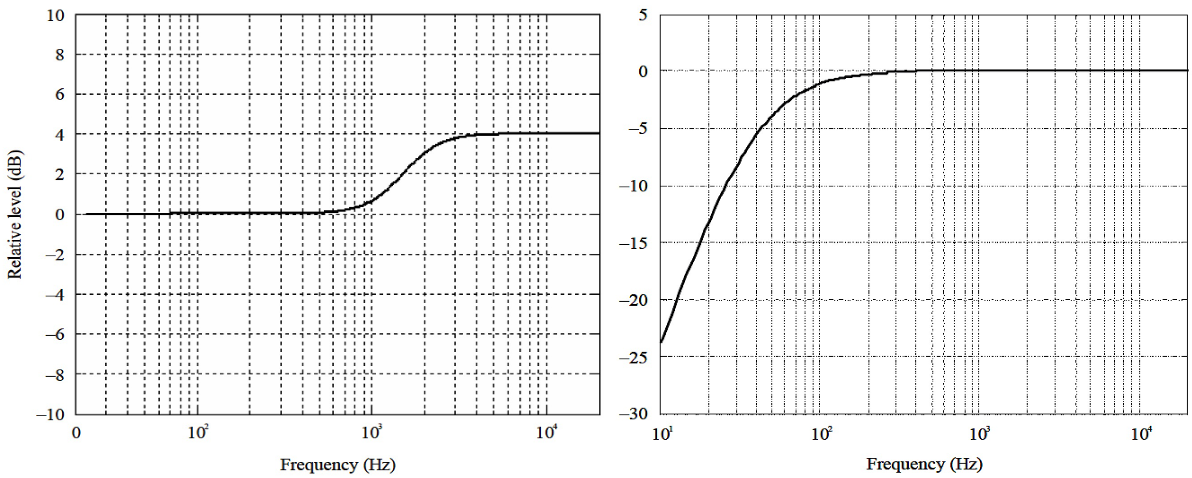


Figure 3.5: Left is the stage one pre-filter, for acoustic effects related to the head; Right is the stage two Revised Low Frequency K and B-Weighted Filters (RLB), a high-pass filter to remove low-frequency energy (ITU-R BS 1770).

Three new measurement units were developed to better measure loudness: LKFS is a unit of measurement designed to reflect usage of RLB filtering. LKFS is the unit designated in ITU BS.1770 standard and (ATSC A/85). However, the EBU uses LUFS, which is identical to LKFS (EBU R128, 2010; EBU R128s1, 2014). LKFS/LUFS (LUFS henceforth) represents absolute loudness level, comparable to the unit digital full scale (dBFS) for the peak level. LUFS maps the measurable parameters of loudness including absolute sound pressure, frequency distribution and duration of the sound (Ronan, Gunes and Reiss, 2017). Loudness Units (LU) is the relative unit used in this system where one LU is equivalent to one dB e.g. LUFS is referenced to 0LU, as dBFS is referenced to 0dBFS.

Finally, three additional measures associated with loudness are created by ITU-R BS 1770: Loudness Range (LRA); True-Peak Level (TPL); and Programme Loudness (PL). LRA replaces dynamic range as a measure of loudness variation within a track. TPL is a measure of the maximum positive or negative measured signal amplitude on a continuous timescale. This measurement may be higher than the largest sample in the digital domain due to the over-sampling (four times in a 48 kHz system to 196 kHz) function of the meter. Systems using a higher sample rate have proportionally reduced oversampling e.g. a 96 kHz system sample rate utilises two times oversampling. PL metering (sometimes referred to as integrated loudness) design provides visual feedback regarding perceived loudness variations between programmes in digital broadcasting and indeed between broadcasting stations (Soulodre and Norcross, 2003; Soulodre, 2004). PL displays an average of the programme material loudness over time. EBU R128 (2011) recommends a target loudness across programmes of -23 LUFS.

3.3 Linear signal processing

Linear signal processing systems must possess homogeneity, additivity and a third property of shift invariance (discussed below). Although shift invariance is not a strict property for linearity, it is a mandatory property for most digital signal possessing techniques (Smith, 1997).

Linearity presented as a graph demonstrates the relationship between the input signal and the output signal known as a *transfer characteristics graph, transfer function* or *input-output graph* (Fig. 3.6). When the ratio between input and output level in a compressor is 1:1, it demonstrates unity gain (Izhaki, 2008).

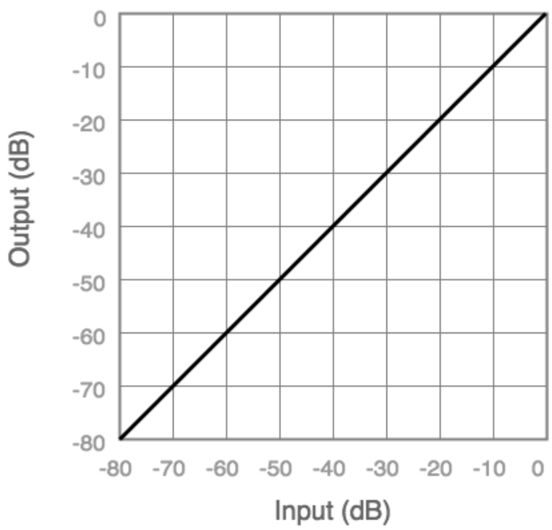


Figure 3.6: Transfer characteristics graph of a linear system having no change in gain from input to output.

3.3.1 Homogeneity

Homogeneity (Fig 3.7) means that a shift in an input signals amplitude results in a corresponding shift in the output signal. Mathematically, if input signal *x(n)* results in an output signal *y(n)*, then input signal *kx(n)* results in an output signal *ky(n)* for any input signal and any constant, *k* (Smith, 1997).

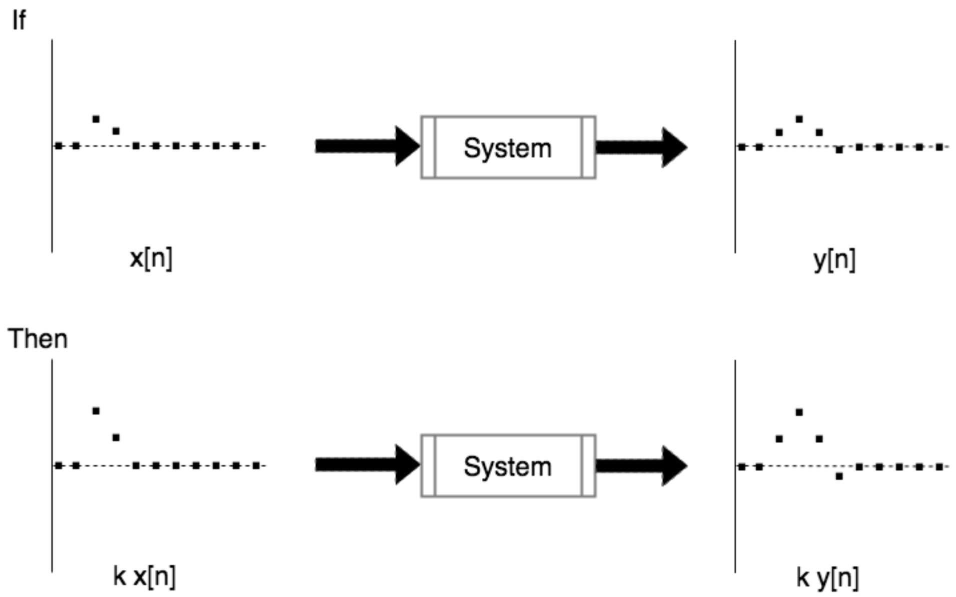


Figure 3.7: The definition of homogeneity (based on Smith, 1997).

3.3.2 Additivity

Additivity (Fig. 3.8) means that added signals pass through the system without interacting (Smith, 1997). Mathematically, if input signal $x_1(n)$ results in an output signal $y_1(n)$ and a different input signal $x_2(n)$ results in an output signal $y_2(n)$, *then t*he system is additive if input signal $x_1(n) + x_2(n)$ results in an output signal $y_1(n) + y_2(n)$ for any input signals (Smith, 1997).

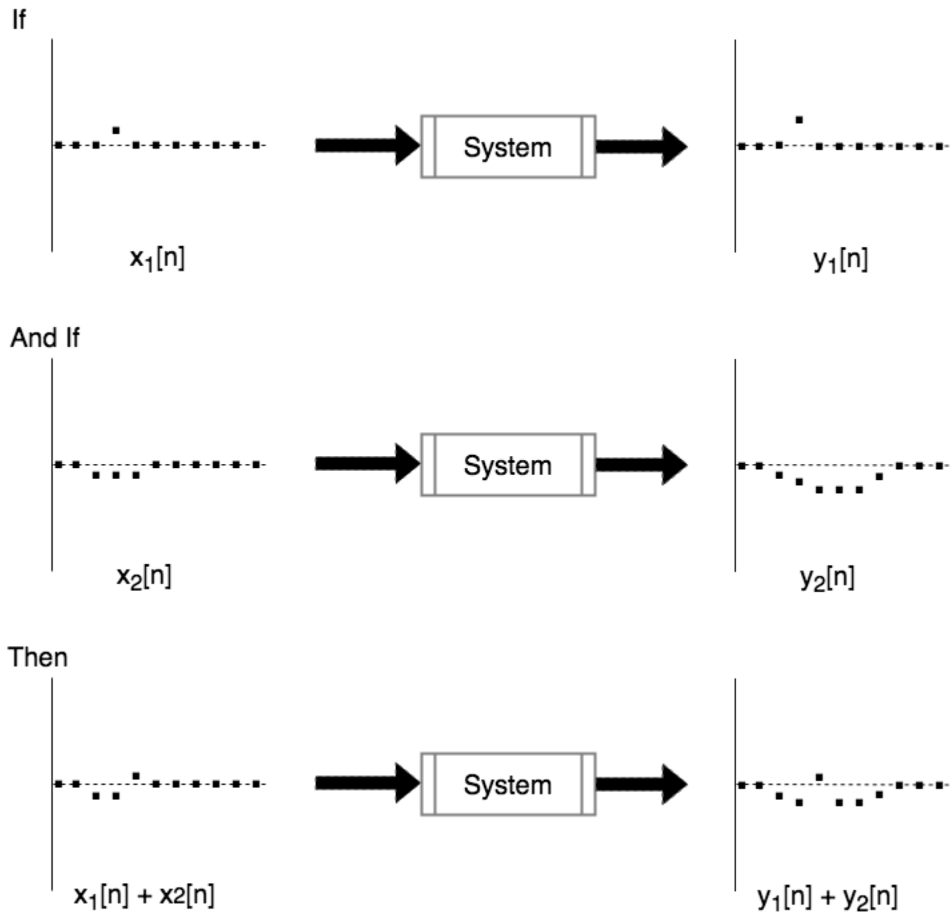


Figure 3.8: The definition of additivity (based onSmith, 1997).

3.3.3 Shift Invariance

Shift invariance (Fig. 3.9) means that a time shift in the input signal results in a corresponding shift to the output signal. Mathematically, if input signal *x(n)* results in an output signal *y(n)*, then input signal *x(n + s)* results in an output signal *y(n + s)* for any input signal and any constant (Smith, 1997).

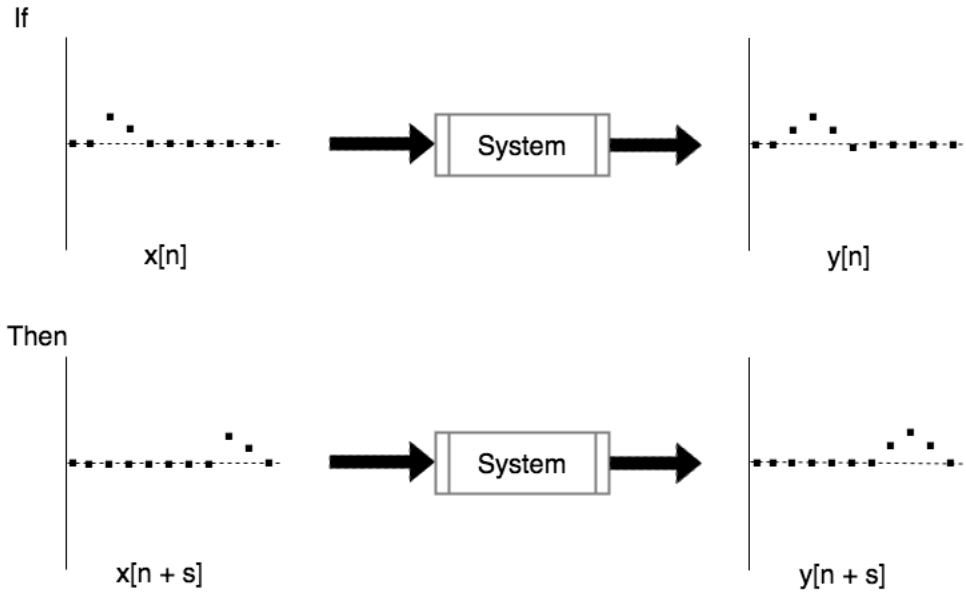


Figure 3.9: The definition of shift invariance (based on Smith, 1997).

3.4 Nonlinear wave processing

DRCs are nonlinear processors that violate the homogeneity and additivity requirements of a linear system. When processing a pure tone sine wave in a nonlinear system, the resultant output signal will display harmonic energy related to the fundamental frequency in addition to the fundamental itself (Fig. 3.10) (Metzler, 2005; Roads, 1979; Zölzer, 2011; Moore et al., 2004). Paraphrasing Roads (1979): 'Nonlinear processing will convert a sine wave into a square wave and will introduce so many modulation products to a complex signal that the input signal itself is blotted out by distortion.' Fig. 3.11 demonstrates the frequency content of a sine wave converted to a square wave, containing only components of odd-integer harmonic frequencies of the original input fundamental. The more the sine wave is clipped, the closer it resembles a square wave, the more evident the odd-integer harmonic components become (i.e. 3(F); 5(F); 7(F); 9(F) and so on).

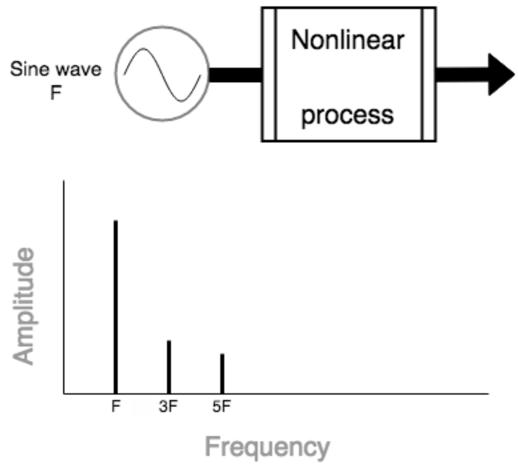


Figure 3.10: The Fast Fourier Transform (FFT) illustrates a nonlinear system and the resultant FFT showing the fundamental frequency of the important sine wave and the additional harmonic components of the input fundamental.

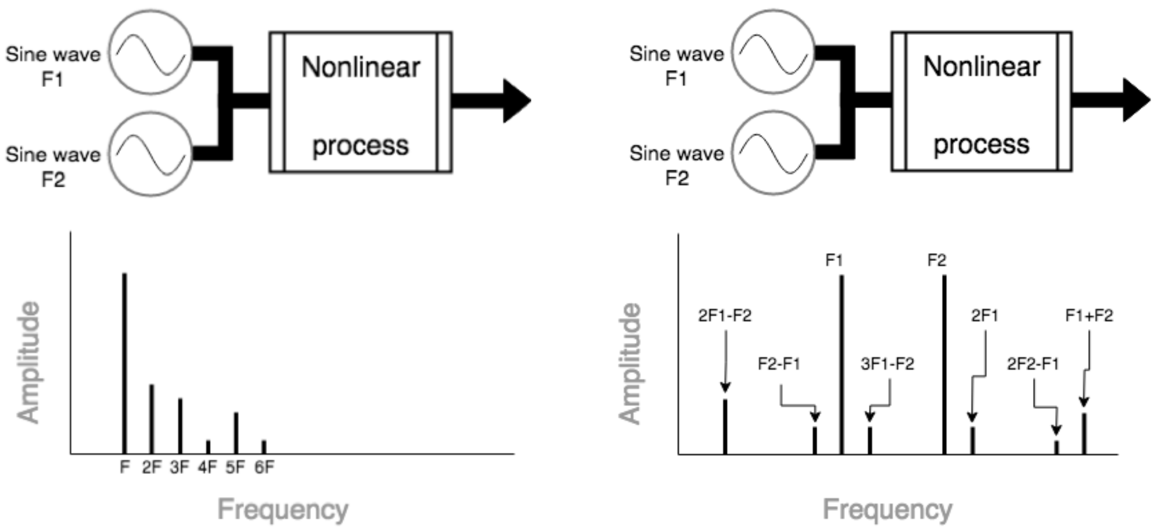


Figure 3.11: The illustration on the left demonstrates two integer ratio signals input into a nonlinear system and the resultant FFT showing the two fundamental frequencies and their harmonic components which are integer ratios of the fundamentals. The illustration on the right demonstrates two non-integer ratio signals input into a nonlinear system and the resultant FFT showing the two fundamental frequencies and their constituent sum and difference components (after Metzler, 2005, pp.26-27).

When processing two modulating signals in a nonlinear system, the resulting output is both fundamentals and each of their harmonic components, as well as IMD (Eq. 13) containing sum and difference frequencies of the two original fundamental signals and their harmonic components (Metzler, 2005) (Fig. 2.20, right) (Roads, 1979). However, if the frequencies in the input are not harmonically related, the spectrum will contain frequencies that "… *cannot simultaneously be overtones of any fundamental, and the output will therefore be in-harmonic*" (Le Brun, 1979). It makes sense that the accumulation of additional frequencies compounds the effect.

Different nonlinear systems produce different types of distortion. For instance, some systems produce nonlinear transfer functions that are rotationally symmetrical around zero (Fig. 3.12), producing odd-harmonic content. An example of this would be the transfer functions of magnetic tape when recording audio at high amplitudes. Nonlinearities which are not symmetrical around zero, such as the transfer function of a section of a semiconductor junction, also produces even harmonics (Metzler, 2005).

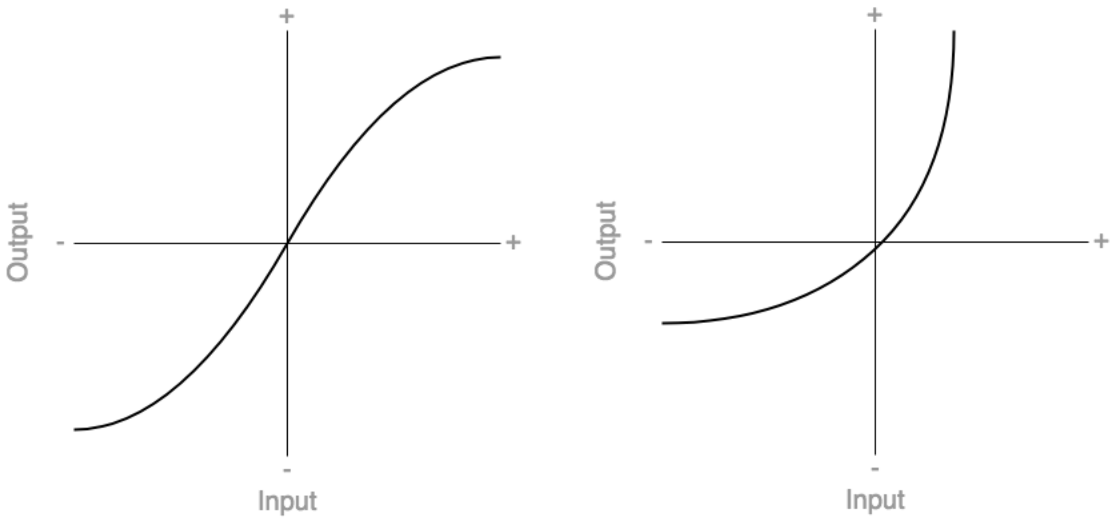


Figure 3.12: Transfer characteristics of two different nonlinear systems. The system on the left (a diode) is symmetrical around zero and the system on the right (a triode) is non-symmetrical around zero (after Metzler, 2005, pp.28-29).

3.4.1 Perception of Distortion

Moore, Tan, Zacharov, and Mattila, (2004) categorise two types of audio signal distortion as a) *linear distortion*, which describes changes in signal amplitudes and the relative phase of frequency components of the original signal, and b) *nonlinear distortion*, which describes the introduction of new frequency components that were not present in the original signal. Moore, Tan, Zacharov and Mattila define the perceptual effects of these distortions: linear distortion is perceived generally as changes in timbre, tonality or 'colouration'; nonlinear distortion results in listeners describing the output signal with levels of 'harshness' or 'roughness' (Moore and Tan, 2003; Tan et al., 2003).

Chapter 4:

DRC Knowledge: How it Works and How it is Used

4.1 Introduction

This chapter discusses the functionality of DRCs. There is a short explanation of musical dynamics with a view for further exploration in the experimental sections. There is also a description of the various dynamic range processors and how they are known to affect signals. Finally, a flow diagram is used to describe and compare the functionality of a typical DRC.

4.2 Dynamics

Dynamics refers to the variations of loudness of a sound, or combination of sounds (Thiemel, 2013). Flat dynamics indicates very little variation between loud and quiet amplitudes; vibrant dynamics indicates plenty of amplitude variation; wild dynamics has excessive variations in the signal amplitude (Izhaki, 2008). The broad classification of dynamics in musical terms is (Table 2.1): ***p*** or *piano*, meaning "soft" and ***f*** or *forte*, "loud". The addition of *mezzo* (half) separates *forte* and *piano*: ***mf*** or *mezzoforte*, "moderately loud" and ***mp*** *mezzopiano,* "moderately soft". Iterations of dynamics notation are: ***ff*** *(fortissimo)*, "very loud" and ***pp*** *(pianissimo)*, "very quiet". In music performance, an increase in loudness over time is *crescendo* and a decrease is called a *diminuendo (decrescendo)*. There are two types of dynamics discussed here, DRC affects both; macro-dynamics and micro-dynamics.

4.2.1 Macro-dynamics

Macro-dynamics describes the variation of signals for events longer than a single note (Izhaki, 2008). Macro-dynamics influence musical expression by creating dynamic variation between musical phrases that form the overall structure of a piece of music. The structure of popular western music utilises combinations of passages: verse, chorus and/or refrain, and bridge and may include an introduction and/or coda (O'Connell, 2001). In musical terms, a scale ranging from pianissimo to fortissimo represent the dynamic intensities within the piece (Deruty and Tardieu, 2014). The macro-dynamics of a musical performance are one factor responsible for conveying the emotion of the music (Juslin and

Laukka, 2003; Bhatara et al., 2011) and perhaps a fundamental characteristic used to make music interesting (Croghan, Arehart and Kates, 2012). Donahue (2008) and Neuhoff, McBeath, and Wanzie, (1999) state that in a natural listening environment, people rarely encounter sounds that do not dynamically change in frequency, intensity, or both. Macro-dynamics aid expression of ideas, to delineate musical ideas and to add drama (Moylan, 2007). In Fig. 4.1, Moylan (2007) illustrates the macro-dynamic programme level dynamic contour of The Beatles' "Here Comes the Sun," (Abbey Road, 1969), illustrating the overall dynamics of the combined instrumentation of the piece of music. In Fig. 4.2, Moylan (2007) illustrates the instrument level macro-dynamic structure for The Beatles' "Lucy in the Sky with Diamonds" (Sgt. Pepper's Lonely Hearts Club Band, 1967). This shows the dynamic variation of each instrument, and the levels of each instrument, as well as the macro-dynamics of the song.

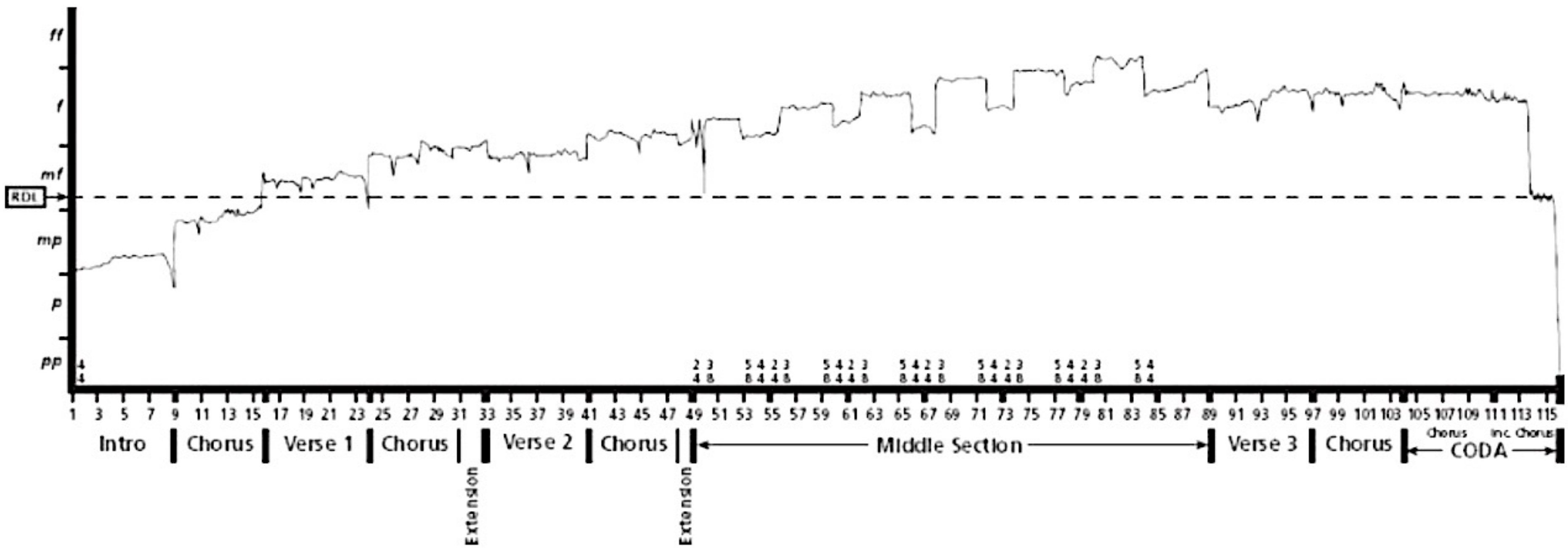
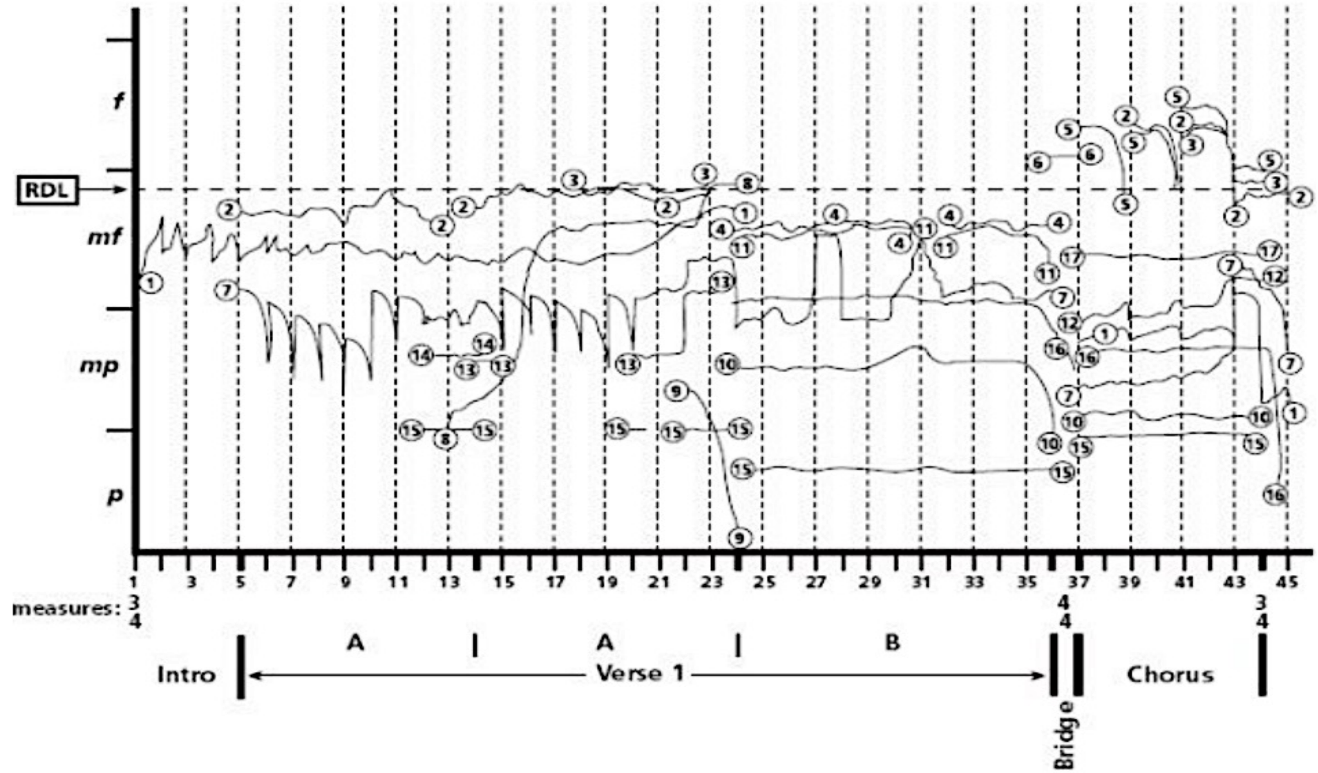


Figure 4.1: Programme dynamic contour of The Beatles' "Here Comes the Sun," (Abbey Road, 1969) (illustration used by permission of William Moylan (Moylan, 2007, p.148).



Key	3 John 2	6 Tom-tom	9 Piano	12 Guitar Bassline	15 Kick
1 Lowry Organ	4 John 3	7 Bass	10 Acc. Guitar	13 Open Hi-Hat	16 Ride
2 John 1	5 Paul 1	8 Tambura	11 Guitar Melody	14 Hi-Hat Closed	17 Snare

Figure 4.2: Instrument dynamics graph for The Beatles' "Lucy in the Sky with Diamonds" (Sgt. Pepper's Lonely-Hearts Club Band, 1967) (compliments o

William Moylan (Moylan, 2007, p.148). Reference Dynamic Level (RDL)

4.2.2 Micro-dynamics

Micro-dynamics relates to the amplitude variations (envelope) within each note. For instance, the envelope of a single snare drum hit would have a rapid attack (transient) and a relatively long decay (Izhaki, 2008) (Fig. 4.3). The attack portion of a signal's envelope is the initial amplitude build-up and settling to sustain level. A piano's attack is the rapid rise in amplitude caused by the hammer striking the strings. The decay portion of the envelope is the amplitude decreasing in level when the excitation is stopped or the instrument is damped and its natural resonances fade to silence. In the case of the violin, after the attack, the note sustains for as long as the player bows the strings and the envelope decays when the bow stops moving and/or removed from the strings.

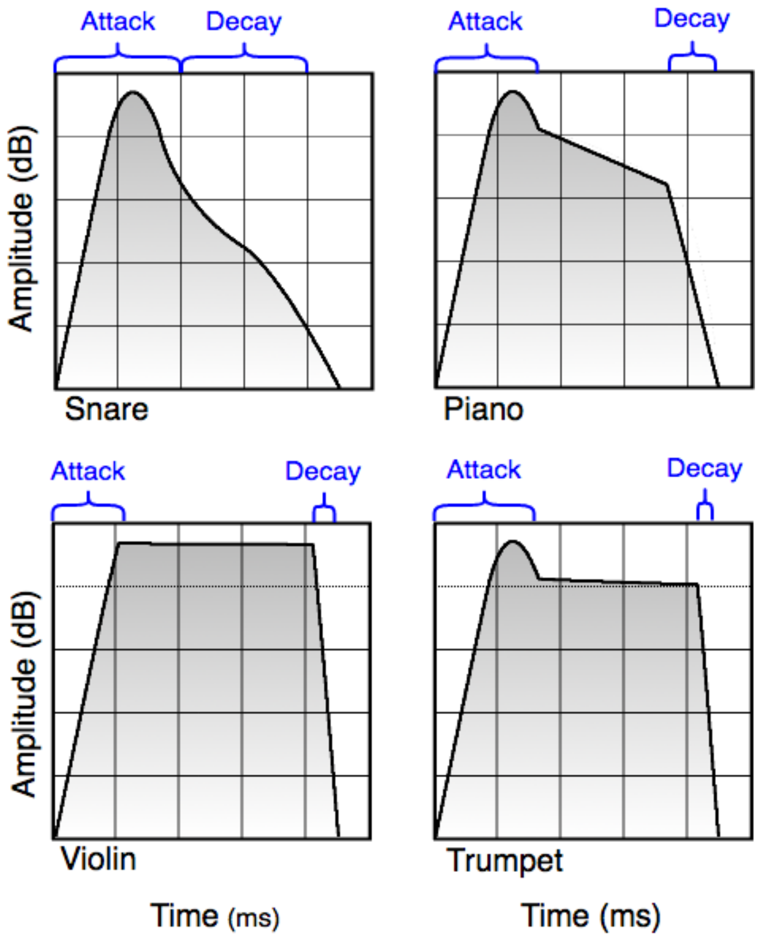


Figure 4.3: Idealised illustrations of the (amplitude vs. time) of four instruments: snare drum, piano, and trumpet (after Izhaki, 2008)

4.3 Dynamic Range Processors

There are several definitions of dynamic range. Dynamic range is the difference in decibels between the inherent noise of a system and the point at which the output signal becomes saturated (system overload) (Young, 1979). Dynamic range for hearing is "*the range of sound level from the smallest audible sound to the largest sound that can be tolerated*." (McGraw-Hill's Access Science Encyclopaedia of Science & Technology Online, 2010). Dynamic range represents the ratio of the largest to the smallest intensity of sound suitable for transmission or reproduction by a sound system. (The Oxford Dictionary of English, 2005) Clearly, quantifying dynamic range of music programmes is not straightforward with no standard definition or method (Boley et al., 2010). This research defines dynamic range using the Audio Engineering Society (AES) definition that the dynamic range of an electronic system is the ratio of the loudest undistorted signal to the quietest undiscernible signal (Bohn, 2000), analogous with the definition of Signal-to-Noise Ratio (SNR). Table 4.1 describes classic dynamic range processors.

Processor Type	Function
Compressor (Downward)	Reduces portion of signal above the threshold.
Limiter	Functions like a compressor, but stops signals exceeding a threshold.
Upward Compressor	Boosts signals below the threshold, making the quieter portions of the signal louder.
Expander (Downward)	Reduces signals below the threshold. Used to maximise signal dynamics. Can increase the signal-to-noise ratio.
Upward Expander	Boosts signal levels above the threshold, raising amplitude of the louder portions of the signal. Can increase the signal-to-noise ratio. These are not generally used.
Gate	Attenuate a signal below the threshold by a specified amount.
Ducker	A form of gate that attenuates one signal dependent on the level of a second signal.

Table 4.1: Dynamic signal processors and their function.

Dynamic range processors are used to manipulate signals with expansion or compression of the signal below or above a specified threshold, respectively. Fig. 4.4 shows the typical transfer characteristics of various dynamic range processors, including idealised graphical representations of input and output signal levels.

Compressors, limiters and expanders manipulate the signal according to a ratio dependent on the input level of the signal. Two special-purpose dynamic range processors function slightly differently. Gate-style dynamic range processors reduce the signal above or below a specified threshold by a fixed amount known as a range. Two types of gates are shown in Fig. 4.5.

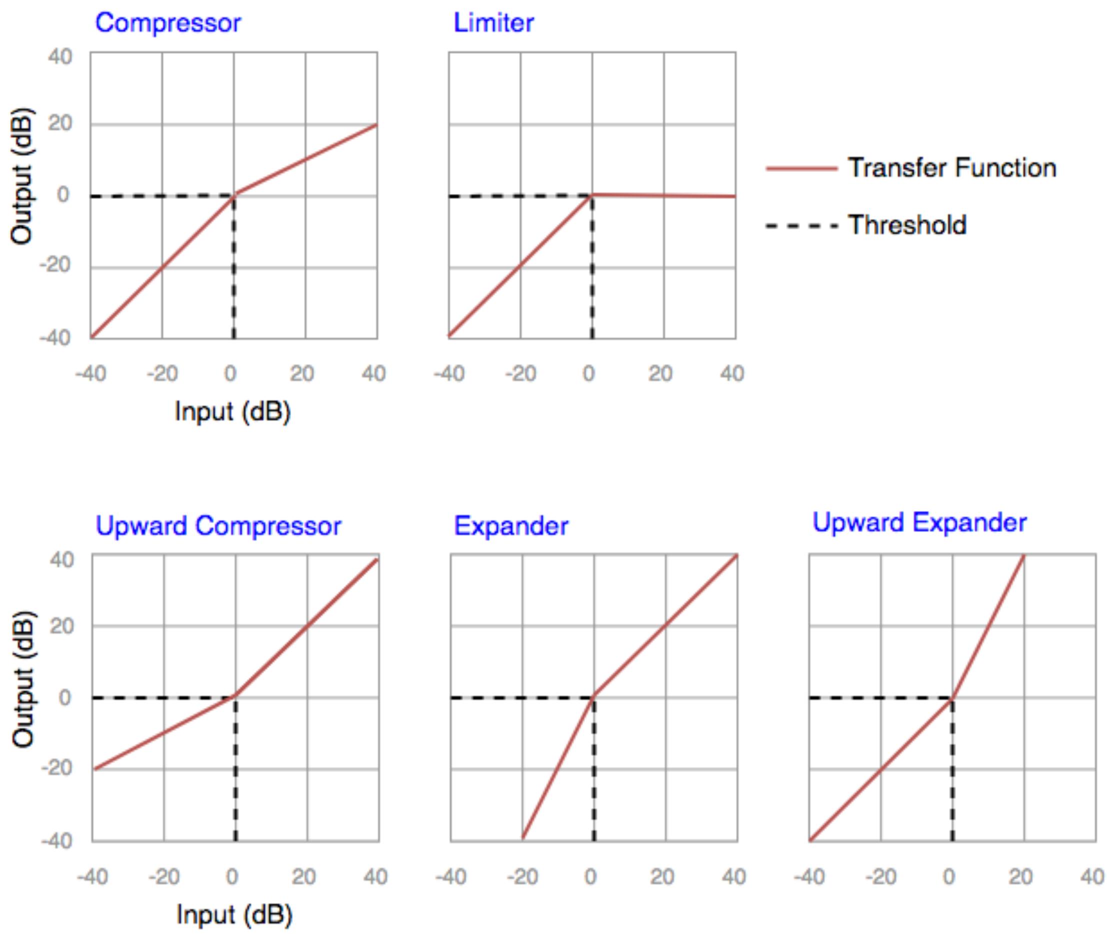


Figure 4.4: Transfer characteristics of typical dynamic range processors (after Izhaki, 2008).

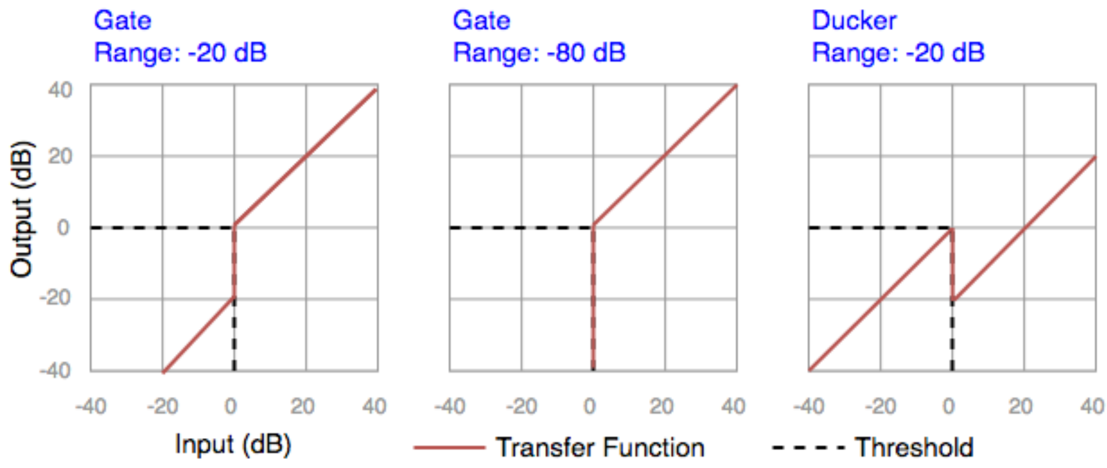


Figure 4.5: Transfer characteristics of gate-type dynamic range processors. (after Izhaki, 2008).

This research focuses on compressors and limiters; therefore, expanders and gates are referenced very little in this thesis. Additionally, downward compression is the predominant design used in DRC and therefore focused on in this research; henceforth 'downward' specification of DRC is omitted, as is discussion of upward compression.

4.3.1 Compressors

Digital compressors and analogue compressors are essentially the same thing, both designed to perform the same task of controlling the dynamic variation of a given signal. Analogue compressors function using analogue components that, by their nature, lack precision and therefore impart unique qualities, even between identical models. The unique qualities of vintage analogue signal processors mean they are particularly sought after in the music industry (Izhaki, 2010; Owsinski, 2008; Owsinski, 2009; Massey, 2009). Some digital compressor designs emulate the inaccuracies found in their analogue counterparts (Izhaki, 2010). An in-depth analysis of the mathematical process of digital compressor design and emulation of analogue circuitry is given in *'Digital Dynamic Range Compressor Design - A Tutorial and Analysis'* (Giannoulis, Massberg and Reiss, 2012).

Compression reduces the level of the signal above the chosen threshold, making the loud sounds quieter. The application of signal gain increases the average signal level and the compressed signal returned to line level. The term 'levelling amplifier' was the popular name for compressors in the 1950s and 1960s, and refers to the principal use of DRC, to level out and control the dynamic variations of the signal.

A variation of compressor design is the multiband compressor; a cross between an equaliser and compressor. Each frequency band of the equaliser (usually three or four) has an independent, configurable compression circuit. Each of the equaliser crossover points is configurable, making it possible to configure the frequency bandwidths to suit the desired application. Multiband compression allows different compression parameters to be applied to potentially problematic spectral areas of a track or mix. For example, low-frequency content of a bass guitar and/or bass drum has DRC applied separate from other frequency bands. Multiband DRC is omitted from this research from this point because of the obvious complications introduced by the added equaliser parameters; although this may form the basis of future work.

4.3.2 Limiters

A limiter is a specific form of compressor, using a very high compression ratio (10:1 or greater) (Owsinski, 2005). The difference between compression and limiting is that a compressor reduces the signal above the threshold but still allows the intensity variations, whereas a limiter allows very little (if any) intensity variation above the threshold. To make a compressor a true limiter would require peak sensing operation (defined below), a hard knee (defined below) and zero attack time (achieved via look-ahead functionality, defined below), which allows gain reduction to happen prior to the signal overshooting the threshold (Izhaki, 2010).

4.4 DRC Technology and Control

This section describes the various functional stages of a typical compressor, referenced with the signal flow diagram above (Fig. 2.24) and transfer characteristic at each stage. The signal flow diagram is based on the design of an analogue Voltage Controlled Amplifier (VCA), typical of modern analogue and digital compressor designs. Compressor and limiter designs are similar and often only differentiated by parameter configuration, therefore, the topography described will not differentiate, unless applicable. From this point DRC refers to the process of dynamic range compression, unless differentiating limiters or compressors is necessary; at which point the specific processes will be named.

4.4.1 Signal input

A signal enters the VCA (Fig. 4.5 a) and is split into two identical signals. Part of the signal is sent to the variable-gain amplifier and the other to the sidechain circuit for level detection/control (Floru, 1999; Giannoulis, Massberg and Reiss, 2012; Izhaki, 2008; Case, 2007).

4.4.2 Gain

The gain stage of a DRC circuit attenuates or amplifies the input signal, determined by the level detecting circuit (side-chain). The gain stage depicted here (Fig. 4.6a) is based on a VCA. However, there are several variations of gain circuitry employed (described below) determining the performance characteristics of the DRC. VCA works by attenuating the input signal according to an external sidechain control voltage (Fig. 4.7) (Floru, 1999; Giannoulis, Massberg and Reiss, 2012; Case, 2007). The sidechain circuitry also makes it possible to feed a modified or entirely different signal into the level detector, the functionality employed in ducker circuits (Izhaki, 2008; Case, 2007).

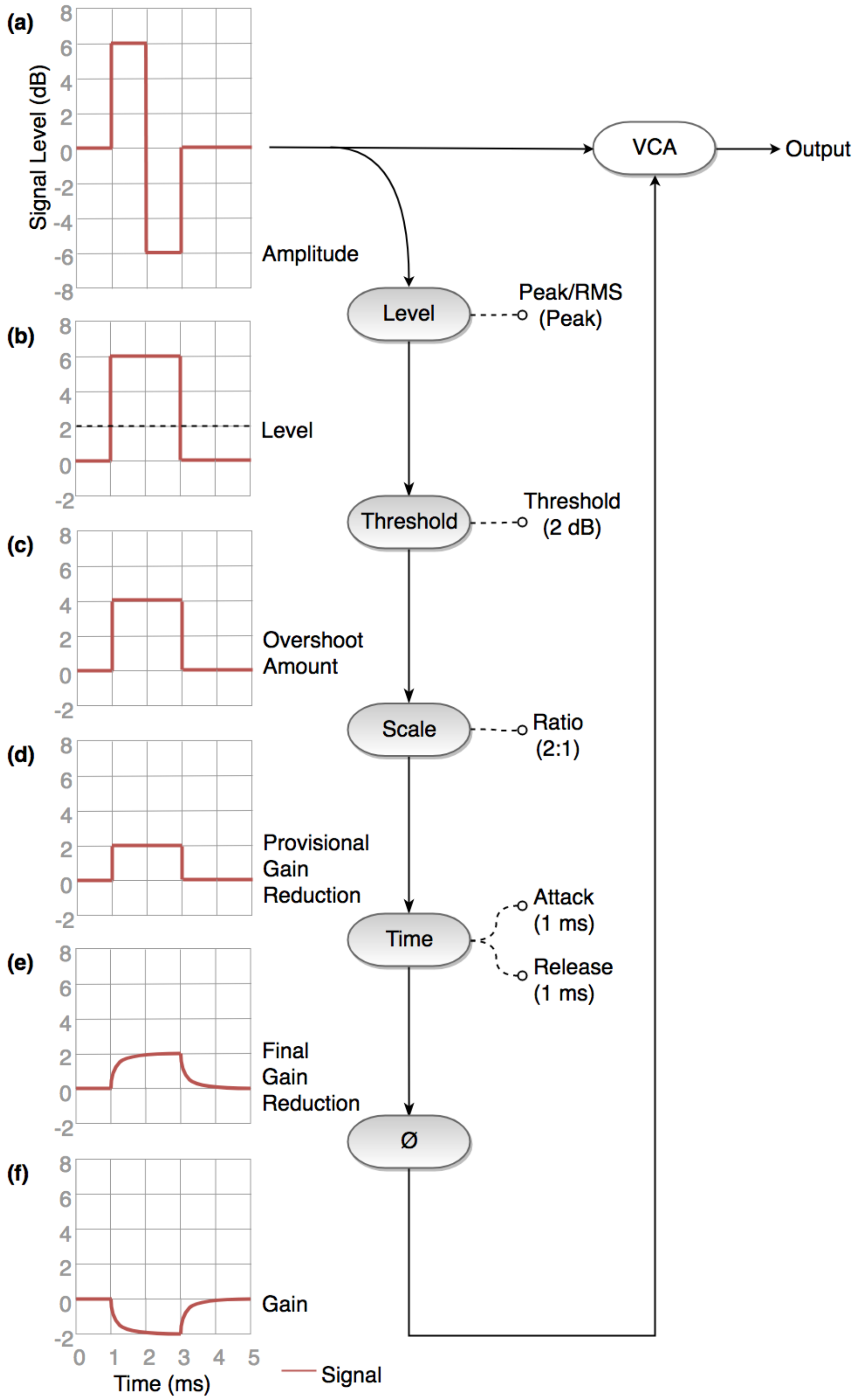


Figure 4.6: Flow diagram describing the topography of a VCA compressor design showing signal flow and controls (after Izhaki, 2008).

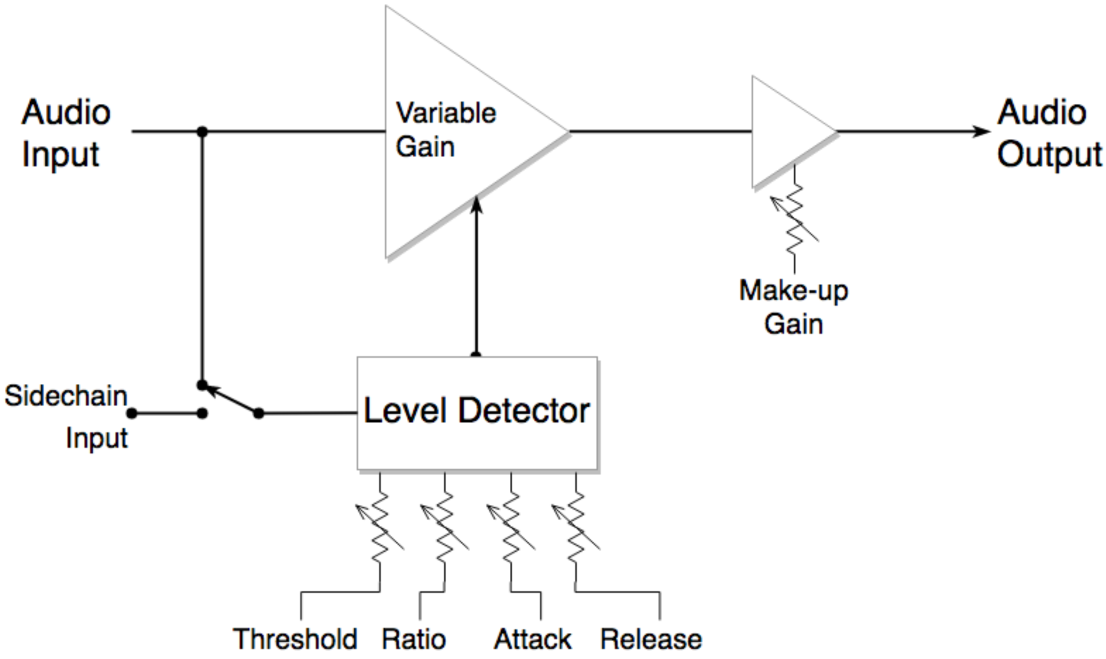


Figure 4.7: Simplified compression circuit (after Case, 2007).

Vari-mu

This is the earliest of DRC design utilising vari-mu vacuum tubes (valves) as gain modification devices (Izhaki, 2008). These DRCs functioned by altering the grid-to-cathode voltage in the tube to change the gain (Giannoulis, Massberg and Reiss, 2012). Vari-mu DRCs have no ratio control parameters and therefore incremental amounts of gain reduction adjust the DRC. After a certain point vari-mu amplification becomes linear, therefore these DRCs have limited gain, a characteristic desirable for use with percussive instruments as the loud transients are largely unaffected. The attack and release operating times are faster than for optical DRCs, but not as fast as VCA or Field Effect Transistor (FET) (Table 4.2) (Tyler, 1979; Garner, 2014; Izhaki, 2008; Case, 2007). Fairchild and Altec built early forms of these DRCs, with a modern state-of-the-art variation manufactured by Manley (Moore, 2013; Case, 2007)

Make and Model	Gain Type	Attack (ms)	Release (ms)
Manley Stereo Variable Mu Limiter/Compressor	Vari-mu	25	200
Universal Audio Model 1176LN	FET	0.2	80
Universal Audio Model Teletronix LA-2A	Optical	10	500
DBX Model 160A Compressor/Limiter	VCA	3	8-400
PreSonus ACP-22 Dynamics Processor	Digital	0.1	50

Table 4.2: Operating times and gain of popular DRCs.

FET

FET replaced tubes in DRCs, utilising much faster attack and release times than the vari-mu, with an added feature of ratio control. FETs return to linearity with very loud input signals, like the vari-mu DRCs (Garner, 2014; Moore, 2013; Giannoulis, Massberg and Reiss, 2012*;* Izhaki, 2008; Case, 2007). An early example is the Urei (now Universal Audio) 1176LN (Owsinski, 2005; Garner, 2014).

Optical

Optical gain DRCs use the interaction of a light source or Light Emitting Diode (LED) and a light-sensitive resistor to affect signal gain. The level-sensitive circuit used in optical compressors is based on the light source increasing in brightness with signal amplitude shining on the photovoltaic cell, increasing the resistance of the photovoltaic cell and therefore increasing the amount of compression of the output signal (Garner, 2014; Giannoulis, Massberg and Reiss, 2012*;* Izhaki, 2008; Case, 2007). Optical DRCs exhibit the slowest response of all DRCs (Izhaki, 2008; Case, 2007). Examples are the Urei (now Universal Audio) LA series (LA-2A, LA-3A and LA-4A) (Owsinski, 2005; Garner, 2014; Case, 2007).

VCA

VCA DRCs are the most popular modern analogue design in current use, employing transistor-based circuitry (Giannoulis, Massberg and Reiss, 2012*;* Izhaki, 2008; Case, 2007). As the name suggests, the detector circuit utilises a control voltage to control gain. DBX employ VCA circuits in their design (Tyler, 1979; Case, 2007).

Digital

Digital compressors employ mathematical computations and are not subject to the same constraints as their analogue counterparts, though there are many examples of digital compressors emulating analogue designs (Izhaki, 2008; Case, 2007). Digital DRCs can also function near-instantaneously, removing any constraints on the attack and release times and offering precise control in their functionality (Izhaki, 2008).

4.4.3 Level detection (Peak and RMS)

The input signal diverted through the sidechain enters the level detection stage (Fig. 4.6b) and converted from a bipolar signal to unipolar. At this point the signal level is analysed by a peak level, average or RMS level detector. The peak level detector senses the maximum level of the signal, estimating the signal level in attack and release phases (discussed below) (Abel and Berners, 2003). The average level detector senses the mean level of the signal, with time detection comparable to a true RMS level detector; both introduce a small time lag due to the feed-back (feed-forward or alternate-feed-back[7]) circuitry (Giannoulis, Massberg and Reiss, 2012). The RMS detector is the only level detector that represents the power of the signal (Floru, 1999). A continuous signal controls the function of RMS sensing DRCs (Katz, 2007) and respond slowly to changes in the input signal, resulting in subtle, transparent compression. Peak sensing DRCs respond

[7] An arbitrary signal is fed into the side-chain to achieve the side-chaining operation as used to implement a ducker or de-esser.

immediately to signal transients resulting in more obtrusive operation (Katz, 2007). The aggressive nature of peak-sensing DRC circuits makes them suitable for limiting applications (Gibson, 2008).

4.4.4 Threshold

The threshold determines the point where gain reduction begins (Fig. 4.6c), depending on the type of dynamics processor, i.e., a DRC attenuates the signal above the threshold (Maddams, Finn and Reiss, 2012), whereas an expander attenuates the signal below the threshold (Izhaki, 2008; Case, 2007). When the signal exceeds the threshold in DRC, called overshoot, attenuation of the signal starts, with increasing attenuation matching the level of overshoot (Fig. 4.8) (Izhaki, 2008). DRC ceases once the input signal falls below the threshold and the processor returns to unity gain (Case, 2007).

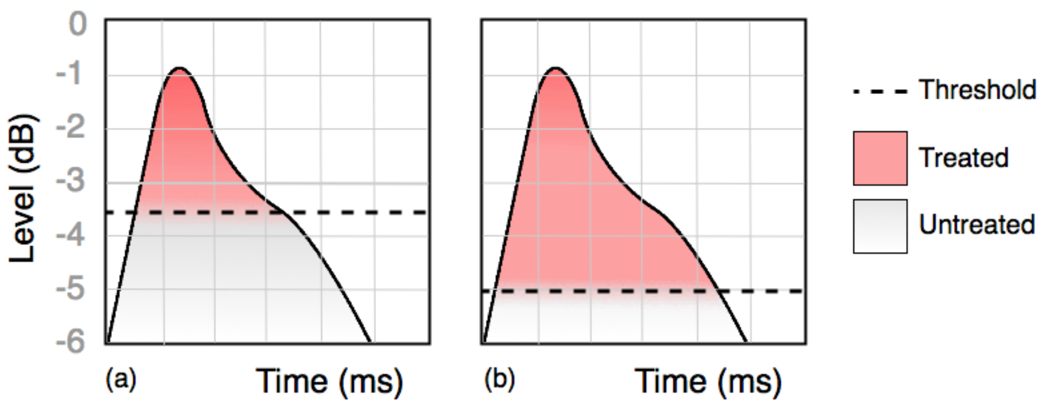


Figure 4.8: Illustrations demonstrating the portions of signals to be attenuated (the treated area in diagram) relative to a high threshold (a) and a low threshold (b) (after Izhaki, 2008).

Compressors either employ a variable threshold or a fixed threshold. The variable threshold employs a dedicated control for adjustment. A fixed threshold means the input gain control determines the magnitude of DRC; increasing input gain equates to added signal overshoot of the threshold and therefore increased DRC, Fixed-threshold DRCs provide an output attenuation potentiometer to compensate for the increased signal level. Fixed-threshold DRCs have the advantage that noise introduced at the input gain stage is attenuated at the output stage; noise introduced at the input stage of variable-threshold DRCs is usually boosted at the output stage when gain compensation is applied (Izhaki, 2008).

4.4.5 Ratio

The ratio setting (Fig. 4.6d) controls the input/output proportion of DRC (Fig. 2.27) (Izhaki, 2008; Giannoulis, Massberg and Reiss, 2012). For example, applying a DRC ratio of 7:1 to a signal exceeding the threshold by 7 dB reduces the signal to 1 dB over the threshold. Similarly, a ratio of 5:1 would result in a 15 dB overshoot reduced to 3 dB (Maddams, Finn and Reiss, 2012; Gibson, 2008). A compressor configured with a ratio of 10:1 or greater will behave like a limiter (Owsinski, 2005). However, a true limiter's function is to ensure that the signal <u>does not overshoot</u> the threshold. For a compressor to behave identically to a limiter it requires configuration with peak sensing level detection, a hard knee (transfer characteristic), and attack time of 0 ms (look-ahead functionality) (Izhaki, 2010). Fig. 4.9 illustrates how the ratio affects a signal exceeding the threshold.

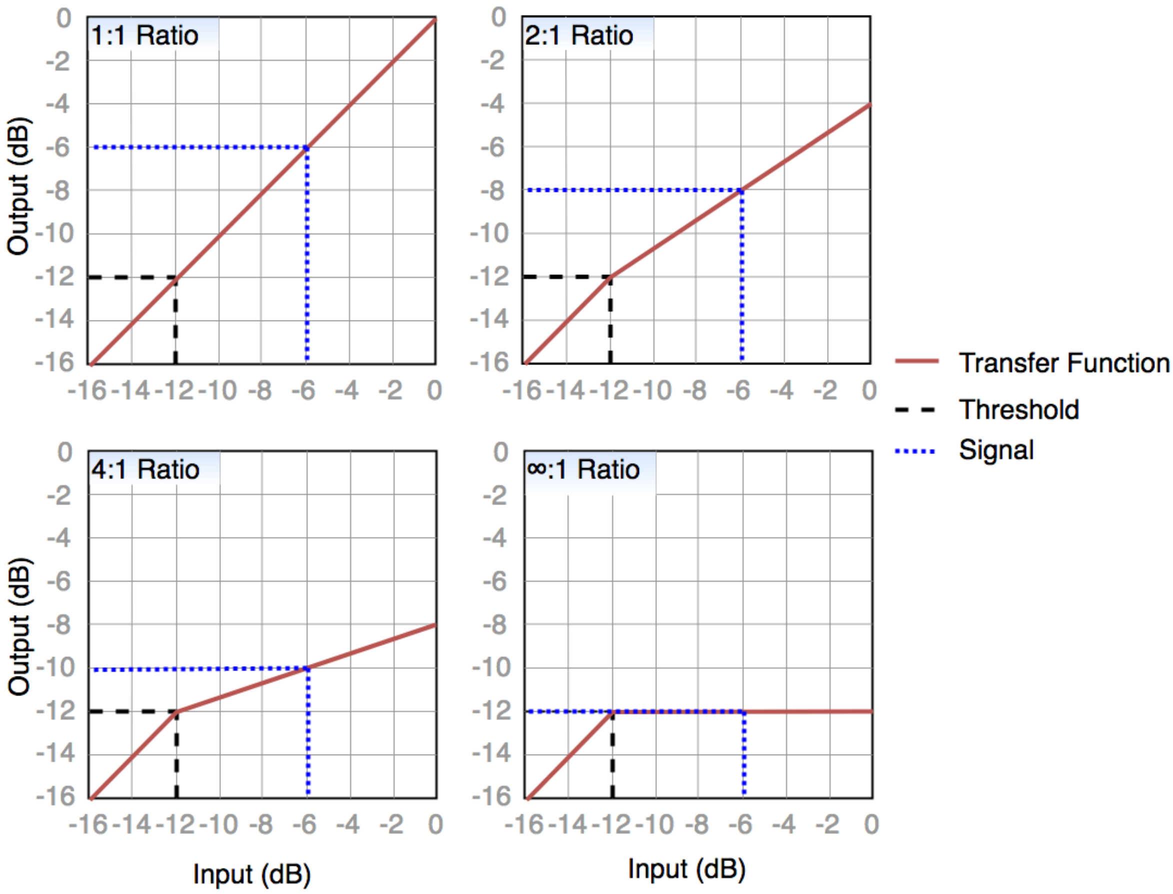


Figure 4.9: Transfer characteristics of unity ratio (top left), 2:1 ratio (top right), 4:1 ratio (bottom left) and limiting (bottom right) (after Izhaki, 2008).

The character of DRC depends on the relationship between threshold and ratio. For example; Scenario 1) lowering the threshold increases DRC, whereas lowering the ratio has the opposite effect. Scenario 2) raising the threshold (reducing DRC) and increasing the ratio (increasing DRC), will accomplish the same amount of DRC as scenario 1, but alter the audible characteristics of the signal (Case, 2007; Izhaki, 2010).

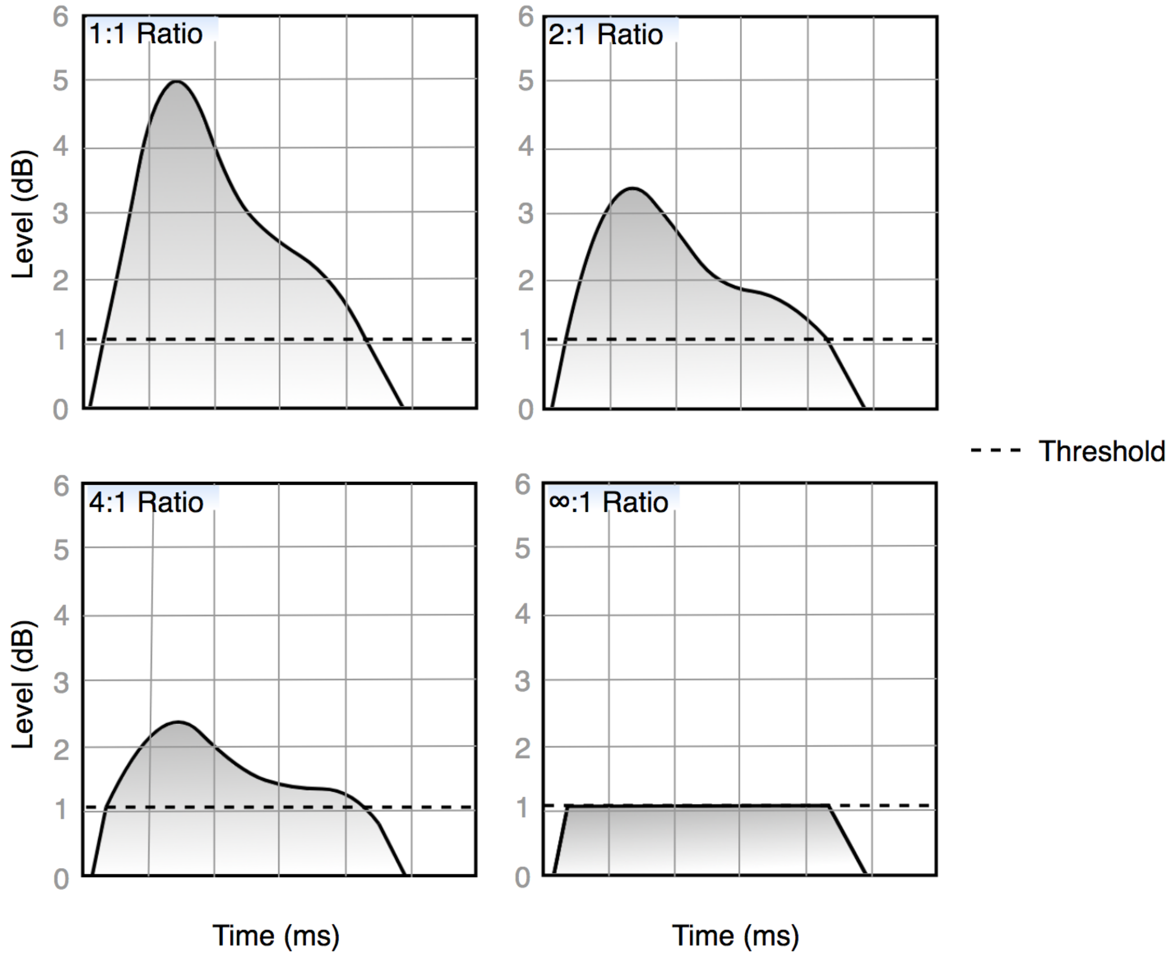


Figure 4.10: Illustration of the effect of DRC ratio on signal attenuation. (after Izhaki, 2008).

4.4.6 Attack and release

The attack function determines how fast the DRC amplifying circuit reacts to an incoming signal (Gibson, 2008) once it exceeds the threshold or indeed enters the knee (Giannoulis, Massberg and Reiss, 2012; Ma et al., 2015). An instantaneous attack (only achievable with look-ahead functionality) will result in immediate amplitude reduction of the signal exceeding the threshold, whilst an exceedingly long attack will result in little to no amplitude reduction, depending on the original signal envelope (Case, 2007; Izhaki, 2010).

Attack and release parameters determine the rate of either increasing gain reduction (attack) or decreasing gain reduction (release). The threshold setting does not influence attack and release function. In other words, should the amplitude of the input signal change while applying DRC, even if change occurs above the threshold, the detected gain changes will initiate the appropriate temporal response of the attack and release parameters.

Fig. 4.11 illustrates the temporal functionality of the attack and release parameters on a signal's envelope. Figure 4.11 (a) illustrates the unaffected original signal envelope; (b) illustrates DRC attenuation only, without attack or release functionality; (c-e) demonstrate increasing attack and release times, and the finite time from activation of DRC to full gain reduction and once the input signal drops below the threshold, for gain reduction to cease (Izhaki, 2008).

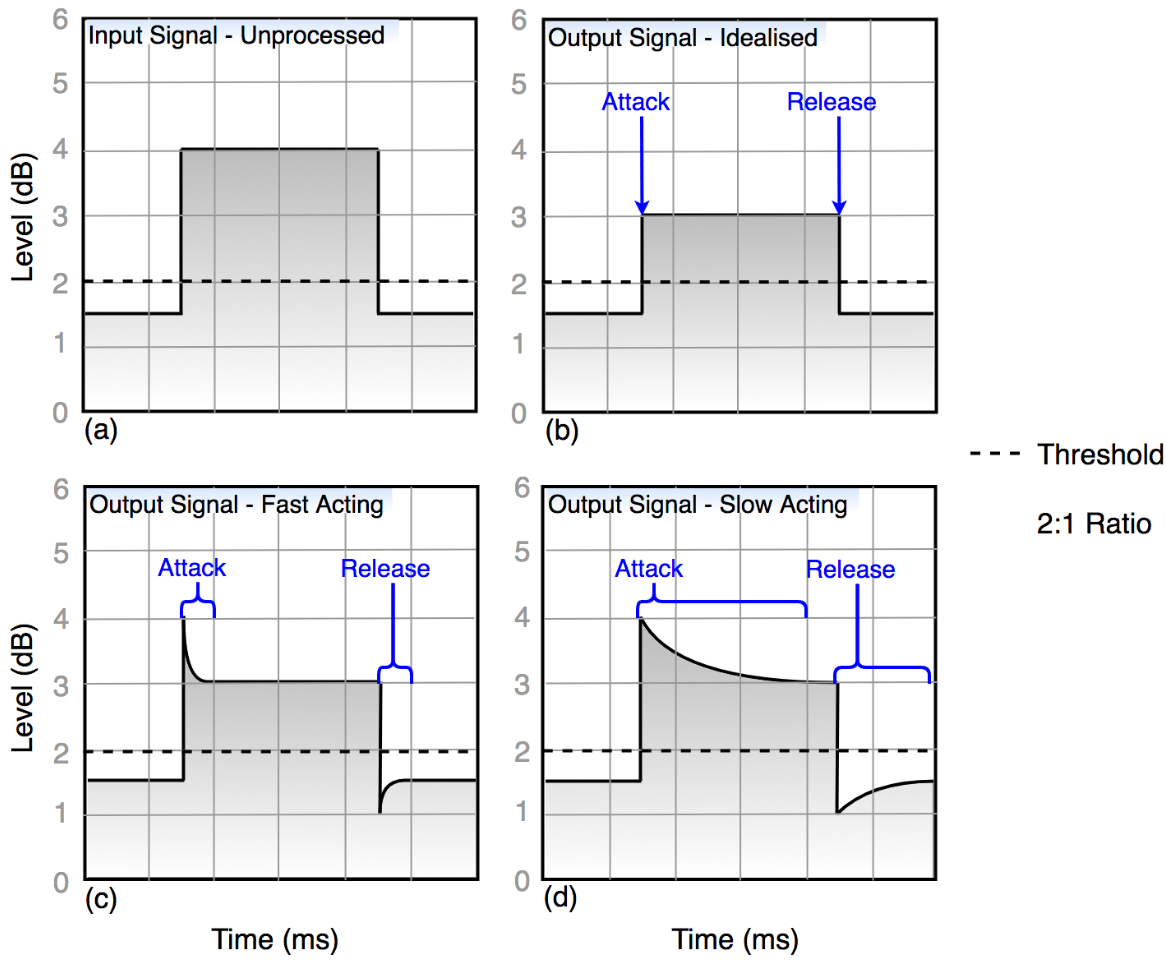


Figure 4.11: The effect of attack and release times on a signal's envelope (redrawn from Izhaki, 2008).

The release time determines how long the amplifying circuit takes to return the output signal to the original level of the input signal once the input signal no longer exceeds the threshold (Giannoulis, Massberg and Reiss, 2012; Maddams, Finn and Reiss, 2012; Ma et al., 2015). Applying too slow a release time can unnaturally modulate the output signal when the input signal is close to the threshold. A short release time can also add audible clicks not associated with the input signal. Short release times can induce unnaturally fast gain recovery of the original signal, which increases the average signal output level. These quick variations in signal level become noticeable variations in signal level, resulting in an audible 'pumping' sound (described in section 2.2.3).

4.4.7 Hard and soft knees

The knee is a parameter that allows adjustment of the transition characteristics between unity gain and the ratio of DRC. A hard knee sets a sharp transitional point. Any signal exceeding the threshold activates the gain reduction and therefore the attack parameter immediately (Fig. 4.12). Upon a signal overshooting the threshold, the output is immediately attenuated by the hard knee, while a soft knee attenuates gradually (Izhaki, 2008).

The immediate nature of the hard knee transition produces a more audible response than the soft knee transition. A soft knee transition activates the gain reduction gradually by allowing gain reduction to start somewhere below the threshold, with a lower ratio, until a point somewhere above the threshold where the selected ratio is reached (Giannoulis, Massberg and Reiss, 2012; Maddams, Finn and Reiss, 2012; Izhaki, 2010; Ma et al., 2015).

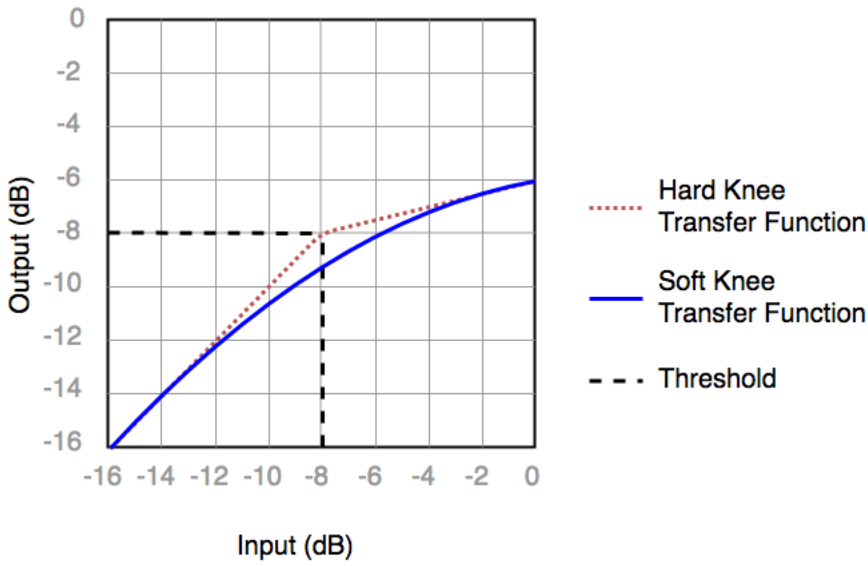


Figure 4.12: Transfer characteristics of hard and soft knee (redrawn from Izhaki, 2008).

4.4.8 Side-chain

Sidechain design can use either feedback or feedforward architecture. Feedback topology means the input to the sidechain is post-gain, making correction of inaccuracies at the gain stage reliable. Early compressors used feedback design. A drawback of feedback design is that it does not allow look-ahead functionality. Feedforward topology means the input to the sidechain is pre-gain. Most modern DRC design is feedforward (Abel and Berners, 2003; Giannoulis, Massberg and Reiss, 2012; Maddams, Finn and Reiss, 2012; Izhaki, 2010; Ma et al., 2015).

External sidechain

Due to DRC design feeding identical input signals to the gain and sidechain stages, the input signal essentially provides level control for itself. Most DRCs permit an external signal feed into the sidechain. The external signal thereby becomes the control signal used to compress the original input signal relative to the external signal's amplitude (Case, 2007; Izhaki, 2010). Sidechain control is what enables the production technique called 'ducking', used especially in radio broadcasting. A normal usage in radio broadcast is to send a music signal into a DRC and feed the announcer's microphone signal into the sidechain; when the announcer speaks the music signal is attenuated (ducked). When the speaking has ceased, the music signal returns to the pre-ducked amplitude (Case, 2007; Izhaki, 2010).

Sidechain filters and inserts

Sidechain filters and inserts allow equalisation of the sidechain signal. Some DRCs provide sidechain insertion points to allow processing by external devices. An example of an application would be compressing drum overhead microphones. The low-frequency (LF) content of drums tends to actuate DRC gain more so than high-frequency cymbals; leading to non-uniform compression of the cymbals from drum-driven gain attenuations. Applying some low-frequency filtering on the sidechain reduces this effect (Izhaki, 2010).

4.4.9 Hold

Hold functionality is available on some but not all DRCs. Hold determines how long gain reduction continues before the release function begins. The hold parameter is a function of DRC temporal control, like attack and release. The design of the hold parameter is to help reduce low-frequency distortion and allow greater reductions to attack and release times (Case, 2007; Izhaki, 2010).

4.4.10 Make-up gain

The nature of DRC usage is attenuation of the input signal. Application of make-up gain returns the processed signal to line level. Compressors use either automatic make-up gain or a manual level control (Giannoulis, Massberg and Reiss, 2012; Ma et al., 2015; Case, 2007; Izhaki, 2010).

4.4.11 Look-ahead

For a DRC to contain the transients of some signals, it needs to be able to deliver a fast response, which is not always possible. Some DRCs have a quick enough response time to affect transients but deliver non-musical results when used in this manner. Look-ahead works by introducing a small delay time (typically 4 ms) to the primary signal so that the sidechain control signal can function slightly in advance of the primary signal. This delay permits the faster attack times required for full transient response and true limiting capability. As with any delay algorithm used for processing audio signals, look-ahead can introduce small phase misalignment issues (Case, 2007; Izhaki, 2010).

4.4.12 Stereo linking

A stereo DRC is essentially two monaural devices combined into one unit, able to function independently, or with the two channels linked as one stereo DRC. When the channels are linked, a single gain control is applied to both channels simultaneously. Configured monaurally the two channels function independently. Some circuit designs used in stereo DRCs use a monaural summation of the left and right channels to feed the left and right sidechains, whereas other designs maintain left and right separation throughout the circuit (Izhaki, 2010; Giannoulis, Massberg and Reiss, 2012).

4.4.13 DRC meters

DRCs typically have three metering options: input, output, and gain reduction. The input meter indicates the signal level at the input stage; the output meter indicates the signal leaving the DRC. The gain reduction meter shows the gain reduction, aiding the adjustment of the various controls. The metering indicates the magnitude of gain reduction, when DRC is engaged or disengaged and the attack and release parameters functioning (Izhaki, 2010).

Chapter 5:

Literature Review of Psychoacoustic and Signal-based Studies of DRC

5.1 Introduction

Two overlapping domains in which DRC research is conducted are speech intelligibility in hearing aid design (see Kates, 2005; Tan and Moore, 2004; van Buuren, Festen, and Houtgast, 1999; Stone and Moore, 2003; Neuman et al., 1998) and controlling the inherent loudness of musical signals (Giannoulis, Massberg and Reiss, 2012), potentially at the expense of audio fidelity (Aarseth, 2012b; Croghan, Arehart and Kates, 2012; Essling, Koenen and Peukert, 2014; Wendl and Lee, 2014; Deruty and Tardieu, 2014; Kirchberger and Russo, 2016; Vickers, 2010; Viney, 2008; Pestana and Reiss, 2014; Gorlow and Reiss, 2013).

This chapter is a review of studies discussing four points of interest relative to the research questions (restated below): **listening preference** (Kirk, 1956; Olson, 1947; Olive, 2012 and 2011; Hjortkjær and Walther-Hansen, 2014); **fatigue** (McGarrigle et al., 2014; Rudner et al., 2012; Kramer, Kapteyn, and Houtgast, 2006; Tulving, 2002; Rudner et al., 2012; Baddeley and Hitch, 1974; Baddeley, 2000; Rabbit, 1968; Sarampalis et al., 2009; Kahneman, 1973; Riley and McGregor, 2012; McGregor et al., 2012; Brown, 1997; Grondin, 2010; Doob, 1971; Fraisse, 1978; Sackett et al., 2010; North and Hargreaves, 1999; Droit-Volet et al., 2004); **audio quality or fidelity** (Croghan, Arehart and Kates, 2012; Kates and Arehart, 2014; Arehart et al., 2011; Ronan et al., 2015; Pestana and Reiss; 2014; Campbell, 2012; Campbell, 2014; Campbell, Paterson, and Toulson, 2014; Toulson, Paterson, and Campbell, 2014; Stone et al., 2009; Croghan, Arehart and Kates, 2012; Neuman et al., 1998; Maddams, Finn, and Reiss, 2012); and **DRC configuration** (Ronan et al., 2015 and 2016; Pestana and Reiss; 2014; Campbell, 2012; Campbell, 2014; Campbell, Paterson, and Toulson, 2014; Toulson, Paterson, and Campbell, 2014; Stone et al., 2009; Croghan, Arehart and Kates, 2012; Neuman et al., 1998; Maddams, Finn, and Reiss, 2012; Giannoulis, Massberg, and Reiss, 2012; Giannoulis, Massberg, and Reiss, 2013; De Man and Reiss, 2013; De Man et al., 2014; De Man and Reiss, 2015; Hafezi and Reiss, 2015; De Man et al., 2015; Ma et al., 2015).

A. Does listening to music subjected to heavy DRC lead to listening fatigue, evidenced by inferior performance in unrelated cognitive tasks? (addressing gap in knowledge *i*).

B. Which DRC configuration leads to the greatest rate of listener preference for a given piece of music? (addressing gap in knowledge *ii*).

C. What is the quantitative impact of the nonlinear properties of DRC on real music signals? (addressing gap in knowledge *iii*).

D. Can DRC configuration manipulation minimise fatigue, maximise listener preferences, and quantitative signal quality? (addressing gap in knowledge *iv*).

5.2 Relevant Preference Studies

Injudicious use of DRC can, according to frequently reported evidence, negatively affect audio quality (Vickers, 2010; Ronan, Sazdov and Ward, 2014). However, the question of whether listeners prefer objectively higher quality audio is not straightforward and may be influenced by both long-term familiarity (which one might refer to as expertise) and short-term training.

In a classic study, Kirk (1956) examined the audio fidelity preferences of 210 college students over a 6-7 week period in the mid-1950s. The students' preferences were measured using A/B/A tests, in which the frequency spectra of five diverse phonograph recordings (string quartet, symphony orchestra, organ popular music and male speech) were altered using four band-pass filters of differing widths (180–3000 Hz, 120–5000 Hz, 90–9000 Hz, 30–15000 Hz), producing six comparisons per stimulus. Surprisingly, Kirk found that, participants *least* preferred the unrestricted (objectively higher quality) recordings (30–15000 Hz), instead preferring a narrower frequency spectrum.

In a follow-up experiment, participants were divided into two groups, with one group invited to thirteen 40-minute listening sessions using unrestricted stimuli like those from the original experiment. The original listening test procedure was repeated, and it was found that the group had altered their preferences and began to select higher fidelity recordings significantly more frequently. The second follow-up group listened to band-limited recordings (180–3000 Hz) and surprisingly then expressed a significant preference for that configuration when the original test procedure was repeated.

Together, these results suggest that untrained listeners who prefer lower-fidelity audio may do so because this matches their prior listening experiences (this study was conducted in the era of low-quality AM-radio), and that listening preferences can be manipulated in either direction (i.e., either towards objectively higher fidelity, or objectively lower fidelity recordings) by habituation (which one might refer to as training, although the preference change educed may be unconscious and not the result of guidance or feedback from the experimenter). It may be that DRC effects that negatively impact *objective* audio quality may be *subjectively* preferred by some listeners, perhaps due to long-term exposure to this process, and raises the possibility that only trained/expert listeners will find DRC-induced distortion disagreeable.

Olson (1947), seemingly in contradiction to Kirk (1956), found that untrained listeners *did* prefer an unrestricted frequency range, but this study related to live musical performances, as opposed to recorded music (Kirk, 1956), which left the question of fidelity preferences in untrained listeners rather inconsistent, and seemingly contingent upon the environment in which the music was heard.

More recently, in Olive (2011, 2012) (additional participants added in latter article), the listening preferences of 18 school children and 40 college students were compared to those of a commercial speaker manufacturer's trained listening panel. The college students comprised pre-trained (expert) and untrained listeners. In experiment 1, four recordings (clapping, female vocal with guitar, female vocal with strings, and female vocal with orchestra) were played at both 128 kbit/s MP3 and CD-quality. Participants completed 12 counterbalanced A/B trials. In each trial, participants heard the MP3 and CD-quality reproductions of each recording four times. Olive found a significant preference for CD-quality over MP3 overall. Around 80% of students with the most listening expertise preferred CD-quality audio, whilst only around 65% of non-expert students expressed this preference. These data also seem to support Kirk's (1956) findings, suggesting that listening preferences are malleable with habituation/training. However, there is also likely a relationship between expertise and general appreciation for audio fidelity, which potentially mediated these results (e.g., it may be that untrained listeners could not hear a noticeable difference or were simply less interested in the music and therefore less attentive during the study).

In experiment 2 of Olive (2011, 2012), two CD-quality commercial popular music recordings, and four different commercial loudspeaker systems were tested, alternating playback from each in a random order until a listener preference was submitted. A preference for the most accurate, frequency-neutral (implicitly higher quality) loudspeaker system was found over loudspeakers with a measured acoustic performance that deviated from an ideal frequency response (implicitly lower quality). Although examining quite different things (audio source quality *vs*. speaker fidelity), these results contrast with Kirk (1956), in which untrained listeners were found to prefer poorer fidelity music reproductions consistent with their prior listening experiences, since participants' prior listening experience in Olive (2012, 2011) was most likely to have been with poorer quality commercial speakers built for personal use (i.e., not professional reference speakers).

Hjortkjær and Walther-Hansen (2014) conducted listening tests that compared the original releases of several popular music recordings with subsequently remastered versions, on the premise that the peak-to-average ratio was smaller on the remasters (implying greater use of DRC). Test stimuli comprised fifteen 15-second clips, each from a different track, along with their remastered counterparts. All stimuli were CD-quality, chosen based on the anecdotal alleged inferior sound quality of the remastered versions. Loudness-equalised clips were played to 22 university music students, all naïve to the purpose of the study. The clips were heard in A/B combinations (master *vs*. remaster), and the participants were asked to indicate their preference. Each A/B pair was presented to subjects twice in the session, with the order of presentation randomised between trials. Subjects were permitted to replay the pairs as many times as necessary to finalise their decision. A two-sided sign test determined that no significant preference for either version was found. However, since the popular musical stimuli used may have been known to participants *a priori*, the interpretation of putative DRC effects may have been complicated by preferences for familiar reproductions of these stimuli. Furthermore, Hjortkjær and Walther-Hansen acknowledge that there were likely to have been other unquantified production differences between each pair of stimuli, such as equalisation and specialisation. The authors report that *"…subjects changed their preference with the presentation of the same stimulus on 46% of the trials*". Their interpretation was an inability of participants to accurately differentiate between stimuli, but it may also relate to insufficient statistical power (the study was based on a relatively small sample size, and the non-parametric sign test (Dixon and Mood, 1946) is known to be underpowered relative to its alternatives), or other limitations in experimental design (see above).

5.3 Fatigue Related Studies

Critics of the 'loudness war' often cite listening fatigue or auditory fatigue as an undesirable product of heavy DRC (Rumsey, 2008; Nielsen and Lund, 2000, 2003; Vickers, 2010; Sherwin, 2007; Smith, 2008). Sound engineers describe listening fatigue as subjective psychological tiring induced when heavy DRC is applied.

> *"By the time you've listened closely (or tried to) to a whole album that's heavily compressed, you end up feeling like Alex at the end of A Clockwork Orange — battered, fatigued by, and disgusted with the music you love." (Devine, 2013)*

Listening fatigue is not the same phenomenon as auditory fatigue described in Section 3.2, though they are potentially related. Fatigue is related to, although not exclusively (Kramer, Kapteyn, and Houtgast, 2006), the reported high level of concentration and effort needed by individuals in order understand conversation in challenging acoustic environments such as cafeterias (McGarrigle et al., 2014). However, research relating to how one should characterise and assess effort or fatigue related to listening is still in its infancy, and whether it is an entirely valid to do so is debated (McGarrigle et al., 2014).

McGarrigle et al. (2014) defined listening effort as: "*the mental exertion required to attend to, and understand, an auditory message*". The Oxford English Dictionary (2006) defines fatigue as: "*extreme tiredness resulting from mental or physical exertion"*. For the purposes of this research, the latter definition retains 'physical exertion' as part of the definition, unlike McGarrigle et al. (2014), as evidence is lacking that the mechanical fatigue associated with the temporary threshold shift (TTS) is associated with fatigue associated with mental exertion.

Rudner et al. (2012) define cognitive capacity as "*an individual's mental resources available for storage and processing of information*". Cognitive capacity encapsulates two related systems of memory: Long-term Memory (LTM) and Working Memory (WM). LTM is a virtually limitless resource of cognitive capacity, including semantic memory (relating to world knowledge) and episodic memory which relates to "*personal knowledge encoded in terms of time and space"* (Tulving, 2002; Rudner et al., 2012). WM constitutes short-term storage and information processing capacity. WM is necessary for such complex tasks as comprehension, learning and reasoning (Baddeley and Hitch, 1974; Baddeley, 2000; Rudner et al., 2012).

There are many ways to measure listening effort, categorised into three broad groups: (1) self-reported subjective ratings such as Multi-Stimulus test with Hidden Reference and Anchor (MUSHRA) (Mason, 2002); (2) cognitive measures of performance dependent on a secondary task such as a perceptual or memory task (McGarrigle et al., 2014); and (3) physiological measures such as pupil dilation or saliva cortisol levels (McGarrigle et al., 2014; Rudner et al., 2012). Testing WM effectively should include elements of both storage and processing and thus comprise primary and secondary tasks (Rudner et al., 2012).

Rabbit (1968) performed three experiments to test speech intelligibility and working memory in the presence of white noise manipulated in intensity relative to voice amplitude.

In Rabbit's first experiment, 36 men and 44 women (aged 26-68 years) participated. Participants listened to 40 lists of eight pre-recorded (spoken) digits in random order, half mixed with noise and half without noise (referred to as clear), at a rate of one digit per second, with 10-second pauses between successive lists. Two experimental conditions were used: in the first, participants were asked to write the digit heard immediately upon hearing it (transcription); in the second, participants were asked to remember the digits for later recall (memory condition). Expectedly, where digits were memorised for later recall, significantly fewer digits were correctly reported than when they were transcribed. When the digits were presented in clear, significantly more digits were correctly reported than when they were presented with white noise. Significant interaction effects revealed that the presence of white noise caused a much more substantial deterioration in the number of digits correctly reported in the memory condition than in the transcription condition, indicating that the distorted signal heard during encoding caused memory to decay at a faster rate, thereby affecting subsequent recall. This demonstrates that hearing poor quality or distracting audio during a cognitive task does indeed appear to negatively impact other cognitive domains, including memory.

In experiment two of Rabbit (1968), 89 men and women (aged 25-69) listened to 56 lists of pre-recorded (spoken) eight-digit sequences in two four-digit groups in random order. Half of the digits were mixed with white noise and the other half were clear. They were presented at a rate of one digit per second, with two-second pauses between groups and 10-second pauses between successive lists. Participants heard both groups of digits, but repeated one group, cued 1.5 s after the end of each list, via the experimenter's instruction "group 1" or "group 2". Findings showed that the rate of recall for the first group of digits was better if the second group of digits was presented in clear rather than with noise, suggesting that hearing noisy stimuli can also impair *previously* remembered items.

In experiment three of Rabbit (1968), 38 men and 86 women (aged 22-68) heard two pre-recorded prose passages of 682 words (passage A) and 712 words (passage B) from Scientific American magazine. Participants listened to each of the two passages in two conditions: 1. with and without white noise over the entire passage; 2. with noise added to the second half of each passage only. After the presentation of each passage, participants completed 10 written questions designed to test their recall of the passages. Five of the questions pertained to the first half and five to the second half. The questions were framed such that answers were unambiguously correct or incorrect. The clear group answered significantly more questions correctly than the split clear/noise. Recall scores for the second halves were also significantly higher for the clear group than for the clear/noise group. Together, these results suggest increased difficulty in recalling verbal information in the presence of noise, even though immediate transcription results showed near-perfect recognition both with and without noise.

Sarampalis et al. (2009) conducted an experiment to address contradictory results in earlier work that proposed that hearing aid users subjectively prefer noise reduction algorithms that are used to process vocal signals, even though these may diminish their ability to interpret speech correctly. Sarampalis et al. used a dual-task paradigm for their tests, a common approach for evaluating cognitive demand (McGarrigle et al., 2014). The dual-task paradigm arises from the theory that a finite pool of cognitive resource is available, but that this pool can be divided between simultaneous tasks (e.g., Baddeley, 1998; Kahneman, 1973). As the cognitive demands of one task increase, so too does the share of cognitive resources used, reducing the availability of mental resources for the competing task. As a result, the demands of one task can be inferred from the change in performance in the competing task (Sarampalis et al., 2009).

In their first experiment, the ability of participants to remember words spoken in a quiet or noisy environment was examined; in their second experiment, the effect of noise level and the use of noise reduction (NR) algorithms on participant’s speed of processing was measured. In experiment one, 25 normal hearing (pure tone thresholds < 20 dB hearing level, *see* Plack, 2010) participants aged 18-26 were asked to listen to sentences from the Speech Perception in Noise dataset (SPIN-R, *see* Bilger et al., 1984). The primary task was to repeat the last word of the sentences; the secondary task was to (later) recall all words.

The stimuli comprised eight lists of 50 sentences, formulated in such a way that half of the sentences contained information that made the last word of the sentence predictable (e.g. "*a chimpanzee is an ape*"), referred to as 'high-context', whilst the other half of the sentences were not easily predicted (e.g. "*she might have discussed the ap*e"), referred to as 'no context'. Stimuli were played either alone (at 65 dB SPL), or in the presence of noise (four people speaking simultaneously in the background). In the noise condition, the noise level was 65 dB SPL and the sentences were adjusted to either -2 or 2 dB SNR. Additionally, in the noise condition, stimuli were either left unprocessed or were processed using the Ephraim-Malah noise reduction (NR) algorithm (Ephraim and Malah, 1984, 1985). Immediately after hearing each sentence, listeners reported the last word they heard, which was recorded by the experimenter. After every eight sentences, a visual cue prompted the listeners to recall the previously reported words (even if they believed them to be incorrect) in any order.

The results of the first experiment suggested that word recall was perfect for both types of sentences (high-context and no-context) in the quiet condition. In the presence of noise, the identification of high-context words was 30% higher than with no-context words. For both word types (high-context and no-context), when SNR was the lowest, performance dropped by approximately 20%. With no-context words, performance was consistently better without NR. With high-context words, performance was similar with or without NR, although at a low SNR performance was slightly better *without* NR. In a three-way repeated measures ANOVA with noise as the dependent variable and SNR, processing, and word type (high-context and no-context) as dependent variables, only the interaction effect between SNR and NR was significant. Four comparisons with and without NR for each SNR and word type (high-context and no-context) revealed no significant differences for NR at 2 dB SNR for either word context. However, with an SNR of -2 dB, performance was significantly better for high-context words and no-context words. Measuring the number of words recalled, a significant three-way interaction between SNR, NR and word type (high-context *vs*. no-context words) was found which the authors interpreted as suggesting that the effect of NR was different for SNR and word type. Four comparisons with and without NR for each SNR and word type showed no significant difference between NR for the two SNRs with no-context words. However, recall of high-context words was significantly better with NR than without NR for a -2dB SNR. Finally, the effect of noise and NR on rehearsal (recalling the words in the order presented or not) was tested with two two-way ANOVAs for each word type (high-context and no-context) as the dependent variables and word position (eight levels) and noise condition (five levels) as factors. The interaction with no-context words was not significant, whereas that for the high-context words was significant. The results were verified using a two-way ANOVA that excluded the quiet condition that found no significant interaction between no-context words but a significant interaction with high-context words. This indicates that NR increases the ability to rehearse high-context words in low SNR environments.

In the second part of their experiment, 25 participants with normal hearing aged between 19 and 27 were asked to repeat speech, heard in quiet or in noise, while performing a simultaneous visual reaction time task. The primary task was to replicate the speech heard in quiet or noise, with a secondary task requiring that they responded as quickly as possible to a computer-displayed visual task. Stimuli comprised 25 Institute of Electrical and Electronics Engineers sentences (IEEE, 1969; recorded by Galvin and Fu at the House Ear Institute), which contained 720 sentences of similar length and grammatical difficulty. The sentences were presented at a rate of 1 sentence every 8 seconds (sentences were approximately 3 seconds in duration), in quiet or in the presence of noise (four people speaking simultaneously in the background) at a level of 65 dB SPL. In the presence of noise (at a level of 65 dB SPL), the volume of the sentences was adjusted to -6, -2 or 2 dB SNR.

Additionally, when the noise was present, stimuli were either processed or unprocessed using the Ephraim-Malah NR algorithm (Ephraim and Malah, 1984, 1985). Participants repeated the sentences, or any part they understood, directly after the presentation, which was recorded by the experimenter. Participants simultaneously performed a computer-displayed visual task (a dual task, unrelated to the auditory stimuli), in which two boxes, measuring 8 cm and 6 cm in height and width, and separated by 1.5 cm, were presented. At quasi-random intervals, a digit [1...8] appeared on either of the boxes. Participants used arrow keys to point an arrow either towards the digit it was even or away from if it was odd. Participants were directed to perform the visual task as quickly and as accurately as possible. The digit remained on the screen until the participant made their direction selection, up to a maximum of 2.5 s. The next digit would appear at a quasi-random interval between 0.5 and 2 s. Accuracy scores and reaction times (RT) were recorded for each trial.

The speech intelligibility component of the task was found to elicit perfect responses from all participants in the quiet condition. In the noise condition, performance decreased to 7% for 2 dB SNR and to 50% for 6 dB SNR, with 4 dB falling somewhere in between. NR had little noticeable effect on performance for the two higher SNRs, but performance decreased by approximately 5% at the lowest SNR. A two-way ANOVA with speech intelligibility as the dependent variable and NR and SNR as factors showed that NR was not significant, but the effect of SNR was significant.

The results of the visual task showed RTs averaging 620 ms for sentences presented in quiet. However, when sentences were presented with noise without NR, RTs slowed by approximately 44 ms for every 4 dB reduction in SNR. With the introduction of NR to the cases with the two highest SNRs, performance was nearly identical to when NR was not used but improved at the lowest SNR with NR. A two-way ANOVA with RT as the dependent variable with SNR and NR as factors showed a non-significant effect for NR and a significant main effect of SNR. In three planned comparisons, there were no significant differences in RT between conditions with or without NR at the two highest SNRs (2 and -2 dB), but at the lowest SNR (-6 dB), RT was significantly better with NR than without. These two experiments indicate that background noise can have a significant negative effect on listening tasks, i.e., that distracting sounds can decrease performance in secondary task performed concurrently that recruits different modalities, thereby impairing apparent cognitive ability.

In Riley and McGregor (2012), 31 children (aged 9-11) with normal hearing were recruited to an experiment that investigated the effect of noise (noise *vs*. quiet) and speech style (plain *vs*. clear) on their ability to learn two sets of eight novel words and their novel object referents (McGregor et al., 2012) presented as line drawings. One half of the referents were combinations of two known animals (e.g., half-pig and-half bird); the other half were combinations of two known inanimate objects (e.g., half-canoe and half-banana). For one half of the children, white noise was mixed with the to-be-learned words at +8 SNR (authors estimated typical classroom SNR; Flexer (2002) recommends a classroom SNR of +15 SNR), whilst the other half listened without the noise added to the to-be-learned words. Additionally, the children heard one set of words in a 'plain speech' style, and another set in a 'clear speech' style. Clear speech was hyper-articulated, producing highly distinctive vowel enunciation (emulating how one might speak in a noisy environment), and tended to be spoken more slowly than the plain speech. Results showed that the children performed at or near perfect for the words presented in quiet. In the noisy condition, children repeated significantly more words correctly under clear speech conditions than in the plain speech conditions, suggesting that noise compromises children's ability to learn new words.

Furthermore, the children's comprehension of the word referents were tested using a four-alternative forced choice (4AFC) task, in which the trained referent was presented with three other referents. Two alternative referents were presented during training, but not labelled (one of which was semantically related to the correct referent) and a third referent never-before-seen that was also semantically related to the correct referent. Results showed a comprehension level at or near perfect across both groups (noise or quiet) and in both conditions (plain or clear speech). Out of a maximum of eight, accurate responses ranged between 7.85 and 7.89. These results seem to suggest that comprehension is not compromised by mild noise, plain speech, or their combination.

Riley and McGregor (2012) then evaluated the number and accuracy of words recalled, using a two 2×2 mixed ANCOVA with the between-subjects factor being the in-noise group (quiet or noise) and the within-subjects factor being speech style (clear speech or plain speech). Participant age and BKB[8]-in-noise scores (clear speech condition) served as covariates. The only significant results from these comparisons were from the quiet group, for which more correct phonemes were elicited than those in the noise group; a marginal result indicated that clear speech elicited more correct phonemes than the plain speech group. These findings seem to corroborate earlier results suggesting that clear speech can ameliorate the effects of noise in learning, and that more demanding tasks require a greater proportion of the available cognitive resources.

Since music rather than speech is the focus of the present thesis, rather than measuring speech interpretation accuracy like the experiments summarised above, the ability of participants to accurately estimate elapsed time is to be measured, which is known to be affected by the presence of auditory and visual stimuli (Brown, 1997; McGarrigle, 2014; Grondin, 2010), despite that, in normal conditions, humans have an in-built mechanism enabling them to accurately estimate the passing of time (Brown, 1997; Danckert and Allman, 2005).

Brown (1997) reviewed over 80 experiments published between 1924 and 1995 to examine the attentional allocation model in which it is proposed that non-temporal tasks disrupt participants' temporal judgements: '*the more attention devoted to time, the greater the lengthening of perceived duration …*' (Brown, 1997). They stipulate that, as noted earlier, attention is thought to be a finite resource, thus it is likely the interference is bidirectional; i.e., that each task receives reduced (likely less than optimal) resources when multiple tasks are undertaken concurrently. Only 9 of the 80 experiments reviewed did not report reductions in perceived elapsed time and/or increased error in time judgments. Brown

[8] Bamford-Kowal-Bench Revised (BKB) Standard Sentence Test (Bench and Bamford, 1979)

applied the review findings to three new experiments that examined interference effects when participants were asked to perform demanding non-temporal working memory tasks, implemented over two sessions, comprising four continuous two-minute self-paced trials to minimise the effects of fatigue (Table 5.1). The experimental design included single-task and dual-task scenarios in conjunction with temporal and non-temporal tasks. It is assumed that temporal and non-temporal tasks draw from the same pool of cognitive resources, and therefore the incorporation of single and dual tasks was expected to produce mutual interference due to resource competition.

Session 1: Single-task Conditions	
Task 1	2-sec timing
Task 2	5-sec timing
Task 3	Easy non-temporal task
Task 4	Difficult non-temporal task
Session 2: Dual-task Conditions	
Task 1	2-sec timing + easy non-temporal task
Task 2	2-sec timing + difficult non-temporal task
Task 3	5-sec timing + easy non-temporal task
Task 4	5-sec timing + difficult non-temporal task

Table 5.1: Tasks for sessions 1 and 2, ordered randomly for each session, from Brown (1997).

In session one, a single-task condition designed to obtain a baseline performance measure was run; in session two, a dual-task condition was used that had temporal and non-temporal conditions. The temporal aspect of the experiment consisted of two-second and five-second timing conditions in which participants produced a continuous series of temporal intervals by pressing a response key repeatedly (every two or five seconds), such that a continual measure of timing throughout the task interval was collected. Note that the temporal measurements (that participant’ generated at specified intervals) are known to have an inverse relationship with verbal estimations and reproductions (Doob, 1971, pp.27-29; Fraisse, 1978, pp.215-217), since the latter is associated with a reduction in the number of temporal cues perceived, therefore allowing a relatively longer time to pass before judging the specified interval has elapsed.

The non-temporal (easy and difficult) aspect of Brown's study differed in the three trials, but each was compatible with the ongoing nature of the concurrent timing-task with degrees of difficulty, which were adjustable. The three trials represented three broad areas of study: motor, perceptual, and cognitive. Experiment one (motor) required that participants manually follow a continually moving visual target using a handheld photosensitive stylus. Trial two (perceptual) required participants to scan a matrix of letters for a target letter. Experiment three (cognitive) was a mental arithmetic task in which participants were presented with a series of subtraction problems and asked to identify which problems had incorrect answers. Participants in each of the experiments were as follows: Trial 1 tested 31 male and female students, trial 2 tested 33 male and female students, and trial 3 tested 30 male and female students. A 2×3×4 repeated measures ANOVA was used, with the following factors: timing (with two levels: two and five seconds); tracking (with three levels: no tracking, slow tracking, and fast-tracking); and task (with four levels: 1-4). Timing performance evaluation was from the mean temporal intervals produced, intended to reflect any overall lengthening or shortening of perceived time intervals, Sx of temporal production score, and the coefficient of variation (COV) each served as (separate) dependent variables.

The simultaneous non-temporal tasks significantly elongated temporal judgments and increased dispersion (Sx) relative to the timing-only conditions. Longer mean temporal intervals signify a shortening of perceived time. Generally, the more demanding non-temporal tasks disrupted temporal productions more than the easier non-temporal tasks. However, the visual tracking and visual search tasks were mostly unaffected by the simultaneous temporal task, with the mental arithmetic task incurring the most significant impairment. Brown's result about the disruption in temporal production performance, produced simultaneously with the non-temporal task, can be interpreted using the attentional allocation model, since fewer resources were available for the temporal component of the activity. However, Brown found no corresponding decline in non-temporal task performance, concluding that if attentional resources are not bidirectional, then the attentional allocation model requires revision to incorporate the idea of specialised independent processing resources.

North and Hargreaves (1999) tested 100 male and female students' estimations of their wait time before the start of an imaginary experiment. Participants were seated in a laboratory and were informed that they could leave at any time. The experimenter left the room, and then started a stopwatch to time how long before participants left the room, up to a maximum of 20 mins, at which point the experiment was halted. One of three types of music were played into the room (subjectively rated on their musical 'complexity'), or a 'no-music' control condition. On leaving the room, or when the experiment was halted because participants did not leave the room, a questionnaire was completed in which participants were asked to retrospectively estimate the length of their wait. Participants perceived their wait time to be significantly shorter in the no-music condition (mean = 15.32 mins) relative to the three music conditions, which differed little: low-complexity (mean = 19.30 mins); moderate-complexity (mean = 18.80 mins); high-complexity (mean = 18.48 mins). North and Hargreaves propose that the music may have distracted participants from their internal timing mechanism. In Cassidy and MacDonald's experiment (2010) participants listened to self-selected music, and estimated their task as having taken longer than when experimenter-selected music had been played. Droit-Volet et al. (2004) suggest that time overestimation may indicate emotional arousal (happy, sad, angry), while time underestimation may be indicative of boredom. It may be that the participants of North and Hargreaves (1999) estimated that less time had elapsed in the no-music condition because of boredom.

5.4 DRC Mix Position Preference Studies

Stone et al. (2009) tested twenty-four normal-hearing university students (12 male, 12 female), aged between 20 and 32 to determine whether DRC impaired performance in a dual-task paradigm. The primary task was to identify and report keywords from two simultaneously presented sentences from the Coordinate Response Measure (CRM) (Bolia et al., 2000) carrier sentence “Ready <CALLSIGN> go to <COLOUR> <NUMBER> now.” Keywords were reported by clicking corresponding scorecard buttons that were assigned to the ‘callsign’, ‘colour’, and ‘number’ options for each of the two presented sentences. During presentation of the two sentences, the scorecards were not available for data entry, and became available following the presentation. On half of the trials participants also performed a secondary task in which a box would appear on the screen as a distractor that that were required to click to activate the scorecard. In all trials, RT was recorded as a measure of cognitive effort. The CRM sentences had varied amounts of DRC applied through alteration of the compression ratio settings: 1:1 (no compression), 1.82:1 (moderate compression), or 10:1 (severe compression). The compression threshold remained constant at 3 dB below the RMS amplitude level of each signal. Finally, the point of DRC application to the two sentences was altered to test the effects of DRC on signals modulating Independently (INDEP) or modulating together (XMOD). For each of the trials, four measures were generated: ‘T_{visdis}’ - time taken to clear the visual distractor; ‘T_{card}’ - time taken to fill the scorecard; ‘Score’ - the number of keywords correct per trial; ‘Reversals’ - when a listener identified keywords correctly (‘callsign’, ‘colour’, or ‘number), but entered them in the “wrong” columns, associated with each of the respective talkers.

Effects related to DRC magnitude for training session				
	NC	***Moderate***	***Heavy***	
A. Interaction of DRC x Visual Distractor for Score (out of 6): $F(2, 46) = 3.72$, $p = 0.032$				
Distractor OFF	4.15	4.22	4.2	NS
Distractor ON	4.13	4.09	3.95	$p = 0.026$
B. Interaction of DRC x Visual Distractor for Reversals: $F(2, 46) = 4.07$, $p = 0.024$				
Distractor OFF	0.373	0.381	0.362	NS
Distractor ON	0.326	0.354	0.401	$p = 0.004$
C. Interaction of DRC x Visual Distractor for $\log10(T_{card})$: $F(2, 46) = 4.41$, $p = 0.018$				
Distractor OFF	0.987	0.989	0.985	NS
Distractor ON	0.953	0.963	0.975	$p < 0.001$
D. Effect of DRC for $\log10(T_{visdis})$: $F(2, 46) = 7.26$, $p = 0.002$				
$\log10(T_{visdis})$	0.134	0.133	0.154	
E. Interaction of DRC x Position of DRC for Score: $F(2, 46) = 3.40$, $p = 0.042$				
INDEP	4.1	4.16	4.12	NS
XMOD	4.18	4.15	4.03	$p = 0.037$
F. Interaction of Compression x Position of DRC for $\log10(T_{visdis})$: $F(2, 46) = 3.22$, $p = 0.049$				
INDEP	0.065	0.06	0.051	NS
XMOD	0.056	0.056	0.062	NS
Effects position of DRC related for training session				
	INDEP	***XMOD***		
G Effect of Position of DRC for Reversals: $F(1, 23) = 7.65$, $p = 0.011$				
Reversals	0.347	0.385		
H. Effect of Position of DRC for $\log10(T_{card})$ during VD trials: $F(1, 23) = 6.97$, $p = 0.015$				
$\log10(T_{card})$	0.864	0.875		

Table 5.2: Interaction effects related to distractor vs. DRC magnitude for training session (A – D), compression position vs. compression magnitude (E – F) and the effects of compression position on reversals (G) and reaction time $\log_{10}(T_{card})$ (H), reproduced from Stone et. al. (2009).

Table 5.2 results are plotted in Fig. 5.1, showing the mean measures for the three sessions (Training, Test 1, Test 2), which found all trials to be significant apart from ‘Reversals’. However, familiarity with the task led to improved performance over time, especially during the training session, with the effects of DRC becoming smaller or disappearing as the test progressed. Therefore, the authors focused mainly on the results of the training session for their statistical analysis (Table 5.2 results E-H plotted in Fig. 5.2). An ANOVA tested each of the four measures above using the following factors: amount of compression (no compression, moderate, severe), position of compression application, INDEP, or XMOD), and with the visual distractor present or absent.

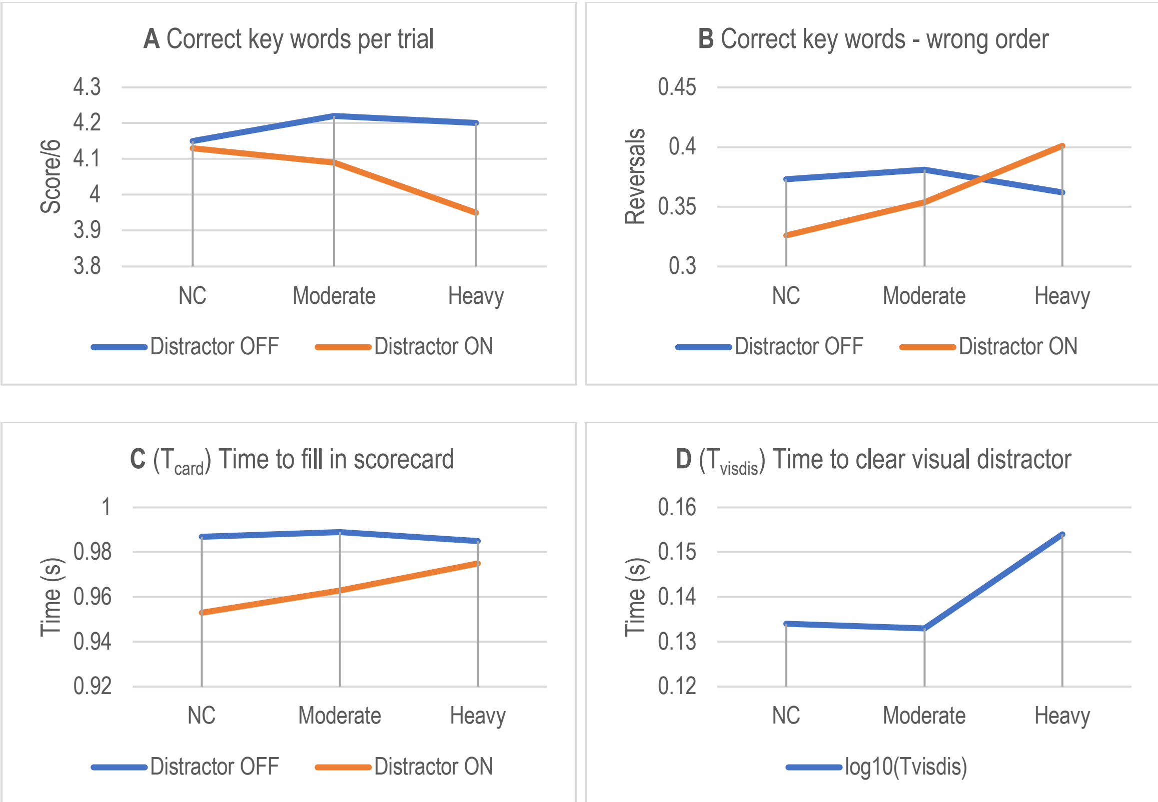


Figure 5.1: Statistically significant effects related to DRC for training session, Table 5.2 A-D results reproduced from Stone et al. (2009).

No significant differences were found for any of the trials without the distractor present. However, the time taken to complete the scorecard increased (i.e., greater RT) when participants heard audio with greater DRC in the presence of the distractor. These results suggest that the compressed stimulus impaired attention/concentration. Importantly, the compression magnitude vs. position of compression (INDEP, XMOD) had a significant effect on participants' score (Table 5.2 E-F plotted in Fig 5.2). Stone et al. (2009) results suggest that although in the INDEP configuration, increased compression magnitude worsened the performance of participants on some measures, when DRC is applied after signal summation (XMOD), the effects are increased.

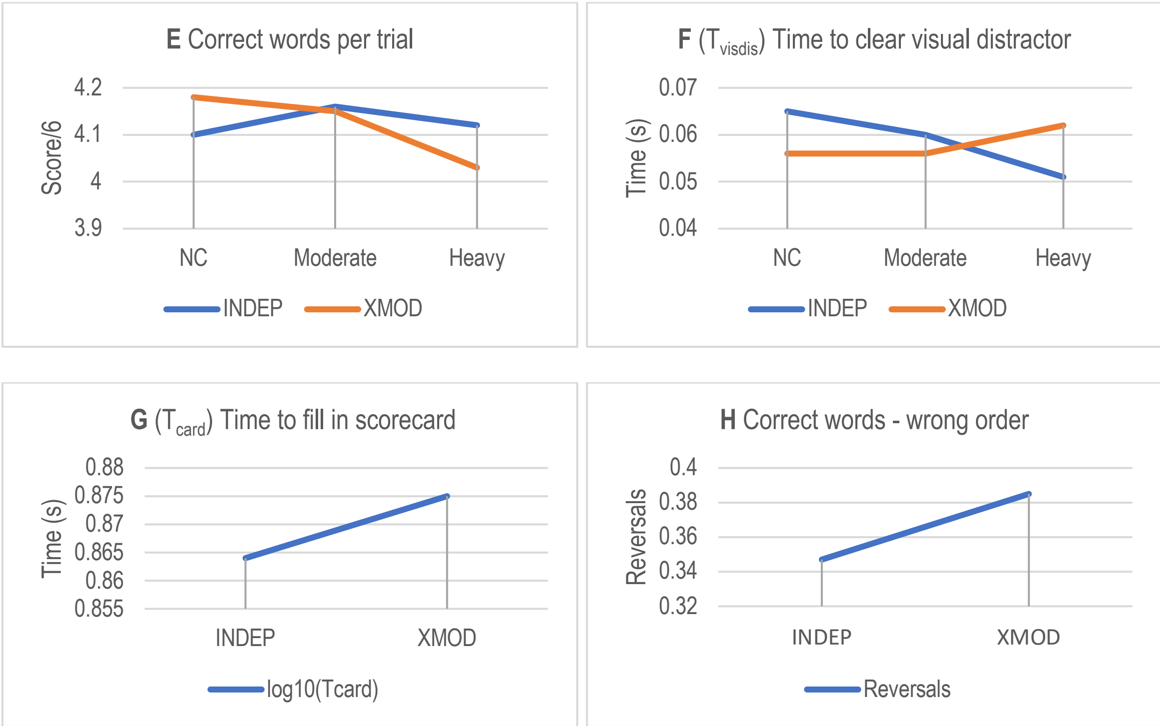


Figure 5.2: Statistically significant effects related to DRC position for training session, reproduced from Stone et al. (2009).

5.5 DRC Quality/Fidelity Perception Studies

Croghan, Arehart and Kates (2012) examined the impact of DRC on perceived audio quality, along with loudness. Two uncompressed 13-second recordings from rock and classical genres served as source stimuli. DRC was applied to the final (summed) signal for each recording, yielding six compression thresholds (uncompressed, –8, –12, –16, –20, and –24 dBFS). Each stimulus had two versions created, one in which loudness was not equalised between stimuli (UNEQ), and another in which loudness was equalised (LEQed). Twenty-three participants rated stimuli on the metrics of preference, loudness, pleasantness, and dynamic range. Participants heard stimuli in randomised pairs, within each genre. Six hours of testing was spread over multiple visits. In both the UNEQ and LEQ conditions, the effect of DRC had a significant effect, for both rock and classical music stimuli, on loudness, dynamic range, pleasantness, and preference ratings (all $p \leq .01$). More specifically, these results indicated a genre-independent preference for light levels DRC over no DRC in the UNEQ condition. The dynamic range ratings for stimuli subject to DRC were markedly lower, showing that the principal outcome of DRC was noticeable. Heavy DRC decreased the likelihood of these stimuli selection as preferred in both UNEQ and LEQ conditions, yielding a commensurate reduction in their pleasantness ratings.

Kates and Arehart (2014) analysed signals using the Hearing Aid Speech Quality Index (HASQI) system, modified to fit the quality ratings for musical signals reported in Arehart et al. (2011). They measured the degree to which DRC affected signal quality by comparing the amplitude envelope and frequency spectra of processed and unprocessed signals, yielding a number between 0 (very low fidelity) and 1 (perfect fidelity). Results indicated a reduction in fidelity with increased DRC; however, the authors overlooked statistical analysis relating HASQI measurements to listener preference, limiting the degree of generalisation from these findings.

Ronan et al., (2015) asked ten award-winning mix engineers to complete a 21-item questionnaire about track subgrouping choices in which questions were formulated using sound engineering literature (Case, 2011; Owsinski, 2013; Izhaki, 2013). Response were examined using thematic analysis (Braun and Clarke, 2006). All participants expressed a preference for the application of DRC to drum and vocal subgroups, assembled to maintain good gain structure (rather than purely for organisation), and stated that pre-DRC subgrouping decisions depended upon musical genre. Similarly, in both Ronan et al., (2015) and Pestana and Reiss (2014), sound engineers/audio mixing students preferred subgroup or full-sum DRC, despite expectations that DRC applied to pre-mixed signal groups to be more deleterious to overall quality because of the amplitude peaks in some signals in a group activating the compressor, thereby affecting other signals within the grouping that would not have otherwise activated DRC.

Using sinusoids, Campbell (2012), Campbell (2014) Campbell, Paterson, and Toulson (2014), Toulson, Paterson, and Campbell (2014) found that applying equivalent amounts of DRC to simple *vs*. summed signals affected the severity and type of distortion introduced. Specifically, reduced nonlinear intermodulation distortion was found when DRC was applied to signals prior to summation, leading the authors to propose that the application of DRC to pre-summed signals may be detrimental to sound fidelity. However, the question of whether findings based upon sinusoidal signals translate to changes in listener preferences for real musical stimuli is unclear.

With such conflict in extant results, and limited consideration of the impact of specific DRC settings on listener preferences for musical stimuli, there is scope for further work to improve understanding of the relationship between listener preferences and DRC. Note that DRC settings are subjective as reflected by the small sample size found in the literature (Table 5.3) including parameters from publications of this author. Particularly with the threshold parameter, anything lower than RMS may or may not be considered severe or heavy DRC e.g. Stone et al. (2009) a threshold set 10 dBFS above RMS is deemed severe DRC while Croghan, Arehart and Kates (2012) discuss heavy DRC yet don't distinguish the threshold between moderate and heavy DRC.

DRC Ratio			
	NC	Compression	Limiting
Stone et al. (2009)	1:1	1.82:1	10:1
Croghan, Arehart and Kates (2012)			∞:1
Neuman et al. (1998)	1:1	1.5:1 2:1 3:1 5:1	10:1
Maddams, Finn, Reiss (2012)			∞:1
Campbell, Toulson, Paterson (2010)			10:1
Campbell (2014)			10:1
Campbell, Paterson, Toulson (2014)			100:1
Toulson, Paterson, Campbell (2014)	1:1	1.1:1 1.2:1 1.4:1 1.6:1 2:1 3:1 4:1 5:1 6:1 8:1	10:1 15:1 20:1
Campbell, Paterson, van der Linde (2017)		1.8:1	10:1

DRC Threshold (dBFS)		
	NC	Light DRC ---> Heavy DRC
Stone et al. (2009)		12 > RMS, 10 > RMS, 3 < RMS
Croghan, Arehart and Kates (2012)	0	-8, -12, -16, -20, -24
Neuman et al. (1998)		20 < RMS
Maddams, Finn, Reiss (2012)		-5, 12 < RMS, -30, -50
Campbell (2014)		RMS
Campbell, Toulson, Paterson, (2010)		-6, -10, -20
Campbell, Paterson, Toulson (2014)		-6, -10, RMS, -20
Toulson, Paterson, Campbell (2014)	0	-5, -10, -15, -20, -25
Campbell, Paterson, van der Linde (2017)	0	-9.9, RMS

Table 5.3: DRC ratio and threshold settings from the literature, including No Compression (NC and 0) parameters.

5.6 DRC Mixing Practice Studies

Research from Queen Mary University developed algorithms over several years to automate and understand music mixing practices by examining such parameters as: audio semantics, equalisation, reverberation, stereo panning, instrument levelling, subgrouping and DRC (Maddams, Finn, and Reiss, 2012; Giannoulis, Massberg, and Reiss, 2012;

Giannoulis, Massberg, and Reiss, 2013; De Man and Reiss, 2013; De Man et al., 2014; De Man and Reiss, 2015; Hafezi and Reiss, 2015; De Man et al., 2015; Ma et al., 2015; Ronan et al., 2015 and 2016; Pestana and Reiss; 2014); Table 5.4 briefly summarises some of the relevant findings. These studies are partially based on interviews and surveys with professional mixing engineers. Pestana (2013) surveyed nearly 60 successful professional sound engineers, who have mixed number one albums or singles, or won a prestigious award for sound engineering. The team has also derived best practice based on practical audio engineering literature (Izhaki,2008; Owsinski, 2006; Coryat, 2008; Gibson, 2005; White, 2000 Case,2011; Case, 2012; Senior, 2012; Katz, 2007). Practice derived from the surveys and literature was implemented and evaluated using subjective listening tests, culminating in DRC algorithms that automate parameters contingent upon side-chain feature extraction (Ma et al., 2015).

Study	Title	DRC Parameters Studied	Brief DRC Findings
Giannoulis, Massberg, & Reiss (2012)	Digital Dynamic Range Compressor Design - A Tutorial and Analysis	Describes several digital DRC designs, including RMS & peak-based, feedforward & feedback, linear & log level detection.	Recommend feedforward DRC as are more stable & predictable than feedback design, smooth performance for wide variety of signals, minimal artefacts, minimal change timbre. Peak detecting DRC is also recommended.
Maddams, Finn, & Reiss (2012)	An Autonomous Method for Multi-Track Dynamic Range Compression	Automation of DRC parameters on tracks of a multi-track mix. Two separate modes defined to automate threshold, ratio and knee.	Ratio may be preferred to threshold for DRC magnitude; Light DRC was preferred to none and heavy; Auto DRC subjectively similar to experienced engineer. Attack & release times automated using spectral flux of signal.
Giannoulis, Massberg, & Reiss (2013)	Parameter Automation in a Dynamic Range Compressor	Automation of DRC attack, release, makeup gain and knee from, side-chain feature extraction of input signal.	Suggests auto make-up gain utilising EBUR-128 and ITU-R BS.1770-2, automate threshold from RMS of signal. Spectral flux used for transient/onset detection for attack and release parameters.
De Man and Reiss (2013)	A knowledge engineered autonomous mixing system.	Auto-mixing system utilising instrument semantics from literature to mix to stereo track using balance, pan, DRC and equalisation	No significant difference between the rule-based (literature derived) autonomous mixing system and pro mixing engineers and outperforms non-semantic automatic system, even though using less sophisticated feature extraction
Pestana and Reiss (2014)	Intelligent audio production strategies informed by best practices.	Literature review and Pro Sound Engineer interviews culminating extensive list of assumptions about mixing, DRC included.	Two main technical reasons for DRC are to control erratic loudness ranges and low-frequency content; Recommend master-buss DRC as best practice; Found the assumption that gentle bus/mix DRC helps 'glue' a mix to be true.
De Man et al. (2014)	An analysis and evaluation of audio features for multitrack music mixtures	Mixing techniques of mixing engineers, focusing predominantly on bass, lead vocal, and drums (specifically kick and snare).	Relative loudness varies by engineer, though average is roughly similar, except vocals - consistently much louder than the rest of the tracks, with a target loudness of about -3 LU
De Man and Reiss (2015)	Analysis of peer reviews in music production.	Critique of mix engineers mixes with analysis of their instrument focus, processing, and ratio of positive and negative comments.	30% of comments on quality and treatment of vocals, 20% on drums and percussion, 11% on DRC automation, DRC and temporary level fixes. 58% on reverb, 16% on panning.
Hafezi and Reiss (2015)	Autonomous multitrack equalization based on masking reduction.	An autonomous multitrack equalization system designed to reduce masking in multitrack mixes	n/a
De Man et al. (2015)	Perceptual evaluation of music mixing practices.	Studies relationship between perceived mix quality & sonic features of mixes.	A distinct preference shown for more dynamic mixes (LEQed) Listeners least preferred overly panned or monophonic tracks.
Ma et al. (2015)	Intelligent Multitrack Dynamic Range Compression	Autonomous multitrack DRC based on side-chain feature extraction. Focuses on automation of DRC attack and release.	The algorithm outperformed or matched performance of semi-professional mixers based on subjective criteria. Listener preference for DRC adjustment is related to style/genre.
Ronan et al. (2015)	The impact of subgrouping practices on the perception of multitrack mixes	Analysis of different subgrouping arrangements and how subgroup processing i.e. Eq, DRC etc.	Drums, guitars, and vocal subgroups are the most popular configurations. Subjective relationship between the number of subgroups used. Subjective preference for subgroups perhaps due to engineer exercising greater control over the entire mix.
Ronan et al. (2016)	Analysis of the subgrouping practices of professional mix engineers	Study of pro mix engineers subgrouping practices.	Subgrouping is to group similar instruments (10/10 respondents) All participants apply DRC at the subgroup stage, most frequently (10/10 respondents) to drum and vocal subgroups. Main reason for subgrouping is to maintain good gain structure (100% median score).

Table 5.4: Studies from Queen Mary University team in development of automated mixing algorithms.

5.7 Guidance on the Experiments to Investigate Research Questions

The studies reviewed above beg the question: does reducing intermodulation of signals under DRC by altering the position of DRC in the mix chain affect listener preference, fatigue, and signal quality?

Studies examining listener preferences for music subject to DRC have typically applied this process to the summed signal (Croghan, Arehart and Kates, 2012; Pestana and Reiss, 2014; Kates, 2005). However, the best point in the signal chain to apply DRC is also the subject of some debate. Using sinusoids, Toulson, Paterson, and Campbell (2014) found that applying equivalent amounts of DRC to simple *vs*. summed signals affected the severity and type of distortion introduced, as reported by Le Brun (1979) and Smith (1997). Campbell, W., Paterson, J. and Toulson, R., (2014) and Campbell, W. (2014) also analysed the impact of altering the point of DRC application on musical signals, derived from observations from the sinusoidal experiments. These preliminary experiments lead to the hypothesis that DRC applied to signals prior to summation reduces intermodulation distortion, leading the authors to propose that the application of DRC to pre-summed signals may be detrimental to sound fidelity. However, the question of whether findings based upon sinusoidal signals translate to changes in listener preferences for musical stimuli is unclear, is explored in this thesis in chapter six.

Rumsey (2008) proposed that 'fatigue' relates to effortful listening and may perhaps be produced by the modern trend of hyper-compressing popular music. Cognitive measures of performance dependent on a secondary task such as the perception of elapsed time while listening to music affected by DRC may indicate fatigue resulting from effortful listening, habituation, stimulus valence/arousal, and general engagement/effort (McGarrigle et al., 2014; Rudner et al., 2012; Kramer, Kapteyn, and Houtgast, 2006; Tulving, 2002; Rudner et al., 2012; Baddeley and Hitch, 1974; Baddeley, 2000; Rabbit, 1968; Sarampalis et al., 2009; Kahneman, 1973; Riley and McGregor, 2012; McGregor et al., 2012; Brown, 1997; Grondin, 2010; Doob, 1971; Fraisse, 1978; Sackett et al., 2010; North and Hargreaves, 1999; Droit-Volet et al., 2004), are tested in this thesis in chapter seven.

Research focused on how and why DRC type nonlinear signal processing affects real music signals is limited. The research of Mahkonen et al. (2013) focuses on de-reverberation and repairing the effects of DRC, using synthesised signals. De Man et al. (2014) use feature extraction to analyse and understand the perceptual evaluation of music mixes, for production and consumption purposes. There is an apparent paucity of research dedicated to understanding the effect of DRC on musical signals that aims to maximise listening enjoyment.

Chapter eight is designed to expand on previous research from the preliminary studies written during and for production of this thesis (Toulson and Campbell 2009; Campbell, Paterson, and Toulson, 2010; Campbell, 2012; Ward and Campbell, 2010; Campbell, 2014; Toulson, Paterson and Campbell, 2014; Campbell, Paterson, and Toulson, 2014. Campbell, Toulson, and Paterson 2010; Campbell, Paterson, van der Linde, 2017). Chapter eight compares the effects of DRC on sinusoidal signals, in various configurations and using the findings to understand the impact of DRC on musical signals better.

Chapter 6:

Listener preferences for alternative dynamic-range-compressed audio configurations

6.1 Introduction

In this study, 130 listeners completed 13 A/B preference trials using pairs of RMS loudness-equalised stimuli subjected to different DRC configurations: *viz*., two magnitudes (heavy, moderate) and two compression types (limiting, compression) applied at three different points in the mix chain (track, subgroup, and master-buss, here termed full-sum), along with an uncompressed control stimulus. The purpose of this study is to ascertain the impact, independently and in combination, of compression magnitude (moderate or heavy), type (compression or limiting), and the point in the mix chain at which DRC was applied (track, subgroup or full-sum).

6.2 Method

6.2.1 Apparatus

Audio stimuli were played on AKG K550 closed-back/over-ear dynamic reference headphones connected to an Apple MacBook Pro running MATLAB with the PsychToolbox/VideoToolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Audio stimuli were created using Avid Pro Tools software that incorporates the peak-sensing Compressor/Limiter III. Stimulus normalisation was performed with SpectraFoo metering software configured to meet the AES Standard RMS reference (AES17-1998). RMS (Eq. 5.1) is representative of the energy or effective power of an AC signal. RMS is used in AC circuits to quantify the root of the average (mean) of the square of a signal's amplitude taken over a specified period. RMS represents the equivalent Direct Current (DC) to produce the same power dissipation over a resistive load (Cabot, 1999; Benson, 1988; Rennie, 2015):

$$X_{rms} = \sqrt{\frac{1}{T}\int_{t_0}^{t_0+T} [x(t)]^2 dt} \qquad \textbf{Eq. 5.1}$$

6.2.2 Participants

130 participants were recruited using opportunity sampling, of which 74 were male and 56 were female (M_{age}= 22.35, Sx_{age} = 5.57). All participants had normal hearing (self-reported). A subset of participants (34) were registered on a cognate (music/audio) undergraduate course, indicating above-average experience with audio, but had no specific training in DRC or audio quality assessment practices. Participants were not trained, except that their task in the experiment was carefully explained (see section 5.2.4), in order that instinctive preference data could be collected, mitigating against 'demand characteristics' (Nichols and Maner, 2008) in which participant responses are biased by perceived experimenter expectations. All participants were naïve to the purpose of the experiment, and were unpaid.

6.2.3 Stimuli

Stimuli were prepared using a 10 s monophonic 16-bit/44.1 kHz recording of a five-piece pop band playing together (as opposed to being overdubbed). The excerpt would not have been known to participants *a priori*, eliminating the potential for any preferences towards a familiar rendition to influence responses (*cf*. Hjortkjær and Walther-Hansen, 2014). The excerpt consisted of eight tracks (1. bass drum, 2. drum overhead, 3. bass guitar, 4. electric guitar, 5. electric piano, 6. viola, 7. vocal 1, 8. vocal 2) without DRC, equalisation, or any other post-production effects, to mitigate against the possibility of unwanted artefacts. The excerpt had considerable amplitude variation (Fig. 6.1), ensuring that, with appropriate settings, the compressor/limiter was periodically activated.

A Pro Tools mix was configured to allow DRC to be applied at one of three points in the signal path (Fig. 6.2): 1] 'Track' (T), having DRC on each track only; 2] 'Subgroup' (S), having DRC on each of three subgroups (consisting of a drum subgroup, vocal subgroup, and other instruments subgroup) only; 3] 'Full-sum' (F), having DRC on the sum of the three subgroups only.

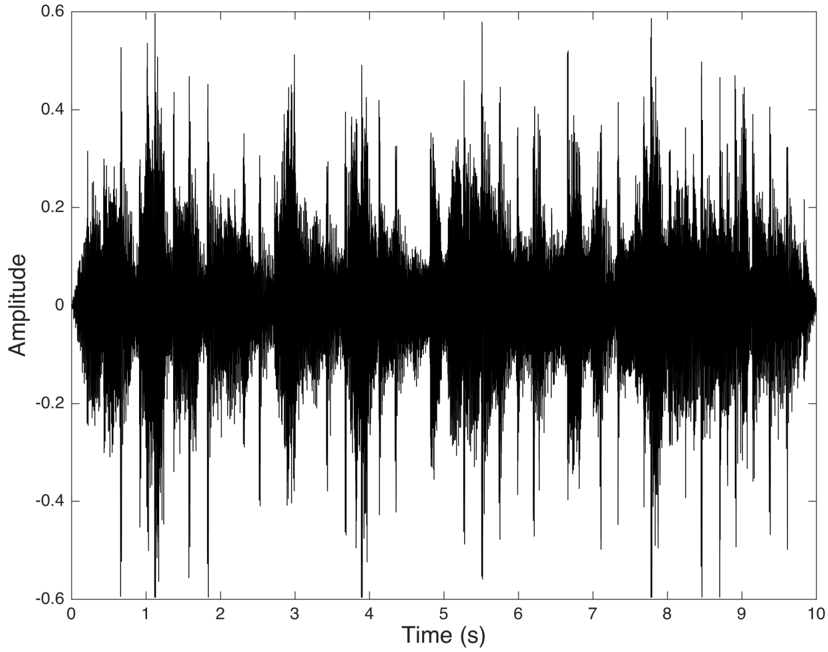


Figure 6.1: Time-domain plot of source stimulus (full-sum).

The Pro Tools Compressor/Limiter III was applied to each track, subgroup, and full-sum simultaneously, in bypass mode as required, ensuring that the signal path was identical for each stimulus generated. The appropriate DRC component was activated as required to produce each of the three compression configurations described above, and deactivated otherwise.

Five combinations of DRC magnitude and type were applied in each of the mix configurations, consisting of Moderate Compression (MC), Heavy Compression (HC), Moderate Limiting (ML) and Heavy Limiting (HL), and No Compression (NC). Differentiation between compression and limiting was by compression ratio, with compression having a ratio of 1.8:1 and limiting having a ratio of 10:1, replicating the settings used by Stone et al. (2009). Differentiation between moderate and heavy compression/limiting was determined by the threshold above which the signal was compressed, with MC defined by 8% of the signal samples exceeding the absolute threshold, and HC by 25% of the signal samples exceeding it, again replicating the compression magnitude settings in Stone et al. (2009).

An algorithm (implemented in MATLAB) was used to reverse engineer the required thresholds for MC and HC by iteratively determining the amplitude thresholds for which either 8% or 25% of the samples exceeded that threshold.

Next, thresholds were multiplied by $20\log_{10}(|amplitude|)$ to convert to standard logarithmic voltage units. This yielded thresholds of -9.9dbFS (8% signal above threshold) for MC and ML, and -14.8dbFS (25% signal above threshold) for HC and HL. The resultant mean absolute amplitudes of all stimuli were within a tolerance of 1%. Stone et al. (2009) used DRC settings of 0.15 ms attack and 5 ms release for spoken voice signals. Here, attack was set to 0.5 ms (with release held at 5 ms), to better suit musical stimuli.

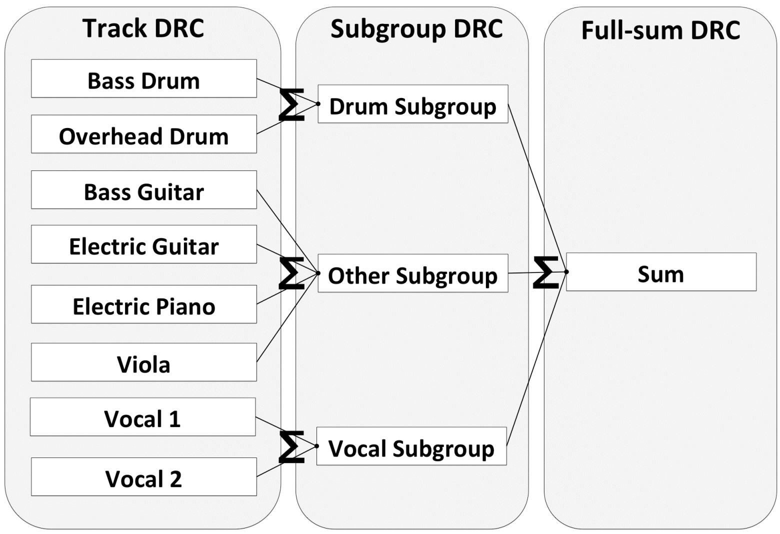


Figure 6.2: Signal path for grouped stimulus production.

A total of 13 stimuli were produced (Table 6.1): Moderate Compression Track (MCT), Heavy Compression Track (MCT), Moderate Limiting Track (MLT), Heavy Limiting Track (MLT), Moderate Compression Subgroup (MCS), Heavy Compression Subgroup (HCS), Moderate Limiting Subgroup (MLS), Heavy Limiting Subgroup (HLS), Moderate Compression Full-sum (MCF), Heavy Compression Full-sum (MCF), Moderate Limiting Full-sum (MLF), Heavy Limiting Full-sum (HLF), and No Compression (NC).

During stimulus production, RMS-normalisation to $-15\mathrm{dbFS}$ was applied at each stage prior to DRC to ensure that signal loudness was maintained (i.e., at track, subgroup, full-sum stages). Additionally, each final stimulus defined in Table 6.1 (including NC) was subject to one final RMS-normalisation to $-15\mathrm{dbFS}$. RMS-loudness, rather than the Loudness Unit (LU) metric specificed in ITU-R BS-1770 (Series, 2011), was used due to the ongoing debate concerning the suitability of LU for musical signals (Pestana, Reiss and Barbosa, 2013; Deruty and Tardieu, 2014), the wide availability of RMS loudness in music production software, and its consequent widespread use by professional sound engineers. However, the LUFS of the processed stimuli used here had a mean of -17.68 LUFS with Sx = 0.09, showing very little variance between stimuli and good agreement (in terms of dispersal) between LU and RMS loudness.

ACRONYM	STIMULUS	DRC CONFIGURATION			DRC SETTINGS			
		Magnitude	Type	Point	Attack (ms)	Release (ms)	Ratio	Threshold
MCT	1	Moderate	Compression	Track	0.5	5	1.8:1	–9.9dBFS (8%)
HCT	2	Heavy			0.5	5	1.8:1	–14.8dBFS (25%)
MLT	3	Moderate	Limiting		0.5	5	10:1	–9.9dBFS (8%)
HLT	4	Heavy			0.5	5	10:1	–14.8dBFS (25%)
MCS	5	Moderate	Compression	Subgroup	0.5	5	1.8:1	–9.9dBFS (8%)
HCS	6	Heavy			0.5	5	1.8:1	–14.8dBFS (25%)
MLS	7	Moderate	Limiting		0.5	5	10:1	–9.9dBFS (8%)
HLS	8	Heavy			0.5	5	10:1	–14.8dBFS (25%)
MCF	9	Moderate	Compression	Full-sum	0.5	5	1.8:1	–9.9dBFS (8%)
HCF	10	Heavy			0.5	5	1.8:1	–14.8dBFS (25%)
MLF	11	Moderate	Limiting		0.5	5	10:1	–9.9dBFS (8%)
HLF	12	Heavy			0.5	5	10:1	–14.8dBFS (25%)
NC	13	No Compression			N/A	N/A	N/A	N/A

Table 6.1: DRC configuration and settings of 13 experimental stimuli.

6.2.4 Procedure

Participants were seated in a secluded, low-noise, low-distraction listening space. Age, gender and audio expertise were recorded. Each participant compared one 10-s stimulus (their *comparator*), selected using the procedure described below, to the 12 remaining 10-s stimuli, and compared the comparator to itself (a control). A sequential, two-interval forced-choice (2IFC) procedure was used (Macmillan and Creelman, 2004), wherein participants were required to choose which audio excerpt they preferred (first or second) by depressing key 1 or 2 on a computer keyboard. A 0.5-s inter-stimulus interval was used, consistent with ITU-R BS.562-3 (Recommendation, I. T. U. R., n.d.). The order that the stimuli were presented to participants was randomised in two ways. First, the order in which the 13 stimuli were presented; second, the order in which the participants' assigned stimulus was presented for comparison to the second stimulus (i.e., first or second). This randomisation procedure was intended to minimise any learning/fatigue/temporal biases caused by stimulus presentation order. Stimuli were intentionally short to enable all stimulus variants to be heard by all participants (all participants completed the experiment in under 5 minutes), and are within the ITU-R BS.562 recommended maximum stimulus duration of 15-20 s (Recommendation, I. T. U. R., n.d.). Each of the 13 2IFC trials began when the participant pressed a key, enabling rest breaks to be taken as desired.

Participant number was used to determine which comparator stimulus was allocated to each participant. For example, participant 1 compared stimulus 1 to each of the 13 stimuli; participant 13 compared stimulus 13 to each of the 13 stimuli. For subsequent participants, the comparator (c) was the modulus of the number of stimuli (n) by participant number (p), i.e. $c = n - p\left\lfloor \frac{n}{p} \right\rfloor$. Since 130 participants performed 13 A/B comparisons, each of the 13 stimuli served as a comparator for 10 participants.

6.2.5 Design

The non-parametric Pearson two-tailed chi-square goodness-of-fit test (Pearson, 1900), shown in Eq. 5.1, was used to determine if observed (*O*) and expected (*E*) listener-preference counts differed significantly. Summing the addend over each of the *i* stimuli in each group of *n* stimuli, it determines if selection rate differed more than one would expect by chance (i.e., at a statistically significant level, with $\alpha = .05$). Expected counts are one half of the number of times that each stimulus was presented (i.e., a per-trial chance selection probability = .5, representing the behaviour of a random responder).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \text{Eq. 5.1}$$

The dependent variable was therefore the frequency with which each stimulus was selected. Independent variables were DRC configuration (with 13 levels), which may be subdivided into magnitude (with 2 levels), point (with 3 levels), and type (with 2 levels), or combinations thereof, along with no compression.

In a closed system such as this, it is the case that if some stimuli are selected less frequently than chance, other *must* be selected more frequently than chance to balance the count. It is therefore difficult to be certain which was the cause and which was the casualty (i.e., did participants particularly dislike some DRC configurations, or particularly like others?), a general limitation of count-based analyses. However, grouped analyses enable us to probe the results further to make reasoned deductions. Therefore, additional combinatorial analyses were conducted, also justified by the postulation of Maddams, Finn and Reiss, (2012), that the interaction of DRC magnitude and type, more so than these effects in isolation, influences perceived audio quality. In these analyses, the independent variable was again stimulus selection frequency, but was contingent upon magnitude and type in combination (pooling over point) yielding 4 levels (MC, HC, ML, HL), type and point in combination (pooling over magnitude) yielding 6 levels (CT, LT, CS, LS, CF, LF), and magnitude and point in combination (pooling over type) yielding 6 levels (MT, HT, MS, HS, MF, HF).

Where chi-square analyses were significant overall, a standard post-hoc pairwise comparison procedure was run. This established whether the observed counts of each stimuli/stimulus group differed from the expected count for that stimulus/stimulus group and the observed and expected counts of all remaining stimuli in that particular analysis, giving one degree of freedom per comparison. Bonferroni correction was used to correct for multiple comparisons, where appropriate.

6.3 Results

Chi-square results are summarised in Table 6.2. Raw listener preference count data is available for download for further analysis.

Independent Variable	Pooling Over	df	N	x^2	p	Significance
Configuration	None	12	1690	64.28	<.001	***
Magnitude	Point & Type	2	1690	3.80	.06	~
Type	Magnitude & Point	2	1690	13.66	.01	**
Point	Magnitude & Type	3	1690	17.37	<.001	***
Magnitude & Type	Point	4	1690	23.97	<.001	***
Type & Point	Magnitude	6	1690	35.21	<.001	***
Magnitude & Point	Type	6	1690	38.46	<.001	***

Table 6.2: Results of Chi-square analyses (~ indicates marginal significance, * indicates $p \leq .05$, ** indicates $p \leq .01$, and *** indicates $p \leq .001$.

The frequency with which the 13 DRC configurations defined in Table 6.1 were selected as preferred was found to differ significantly from chance [$\chi^2(12, N = 1690) = 64.28$, $p < .001$], Fig. 6.3. Post-hoc pairwise comparisons indicate that HCT and HCS were selected more frequently than expected ($p < .01$ and $p = .02$ respectively), whereas HLS and HLF were selected significantly less frequently than expected ($p = .02$ and $p < .001$ respectively). MCT, MLT, HLT, MCS, MLS, MCF, HCF, MLF and NC were not selected at a frequency that different significantly from chance ($p > .05$). Using Bonferroni correction for four comparisons (i.e., each stimuli giving a provisionally significant result *vs.* all remaining stimuli), the threshold for declaring significance is reduced to $p < .0125$, meaning that only HCT and HLF remain significant, with HCS and HLS becoming marginal. However, Bonferroni correction is known to be particularly conservative; subsequent grouped analyses allow effects to be examined using fewer pairwise comparisons.

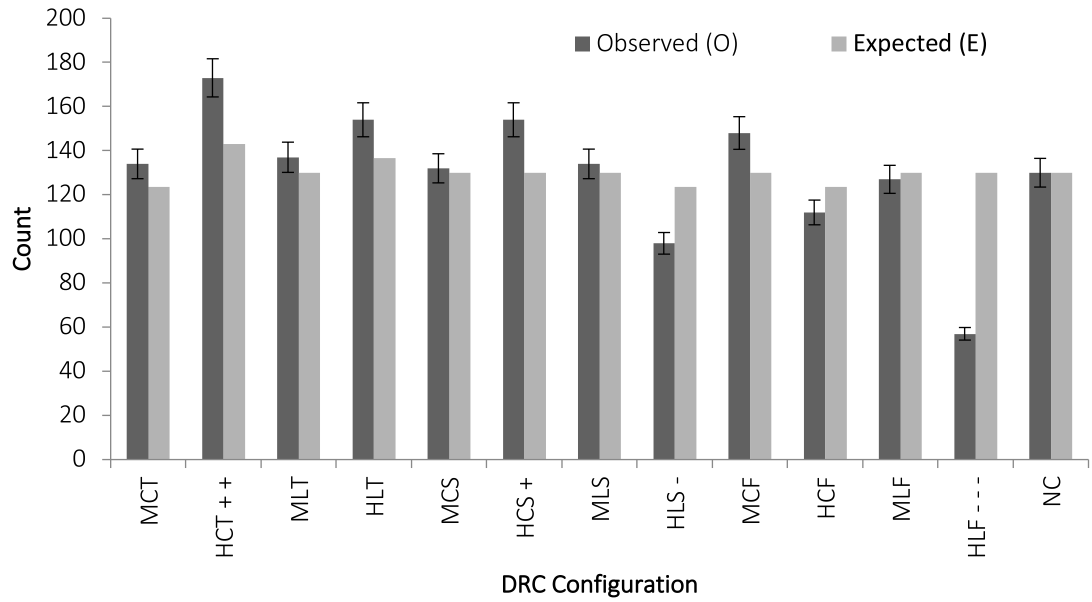


Figure 6.3: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for all DRC configurations. --- indicates observed < expected at $p < .001$, - indicates observed < expected at $p < .05$, + indicates observed > expected at $p < .05$, and ++ indicates observed > expected at $p < .01$. Error bars show ±5% error.

The frequency with which different DRC magnitude settings (moderate and heavy) were selected as preferred was not found to differ significantly from chance [$\chi^2(2, N = 1690) = 3.80, p = .06$], i.e., no significant difference in DRC magnitude preference was observed when data were pooled across other independent variables (type and point). However, this result was marginal (see section 5.3 for discussion), so no post-hoc pairwise comparisons were performed.

The frequency with which different DRC type settings (compression and limiting) were selected as preferred differs significantly from chance [$\chi^2(2, N = 1690) = 13.66, p = .01$], i.e., some DRC type settings were selected significantly more or less frequently than expected when data were pooled across other independent variables (magnitude and point), Fig. 6.4. Post-hoc pairwise comparisons show that Compression (C) was selected significantly more frequently than expected ($p < .001$), whereas Limiting (L) was selected significantly less frequently than expected ($p < .001$). No compression (NC) was not selected at a frequency that differed significantly from chance. These findings survive Bonferroni correction for two comparisons.

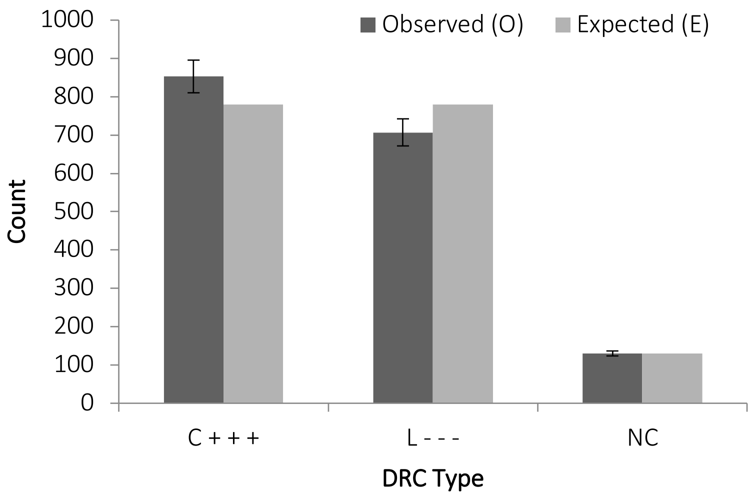


Figure 6.4: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for DRC type settings. (C: compression; L: limiting; NC: no compression), --- indicates observed < expected at $p < .001$, +++ indicates observed > expected at $p < .001$. Error bars show ±5% error.

The frequency with which different DRC point settings (track, subgroup and full-sum) were selected as preferred differs significantly from chance [χ^2 (3, N = 1690) = 17.37, $p < .001$], i.e., some DRC point settings were selected significantly more or less often than expected when data were pooled across other independent variables (magnitude and type), Fig. 6.5. Post-hoc pairwise comparisons show that Track (T) was selected significantly more frequently than expected ($p < .001$), whereas Full-sum (F) was selected significantly less frequently than expected ($p < .001$). Subgroup (S) and No Compression (NC) were not selected at a frequency that differed significantly from chance. These findings survive Bonferroni correction for two comparisons.

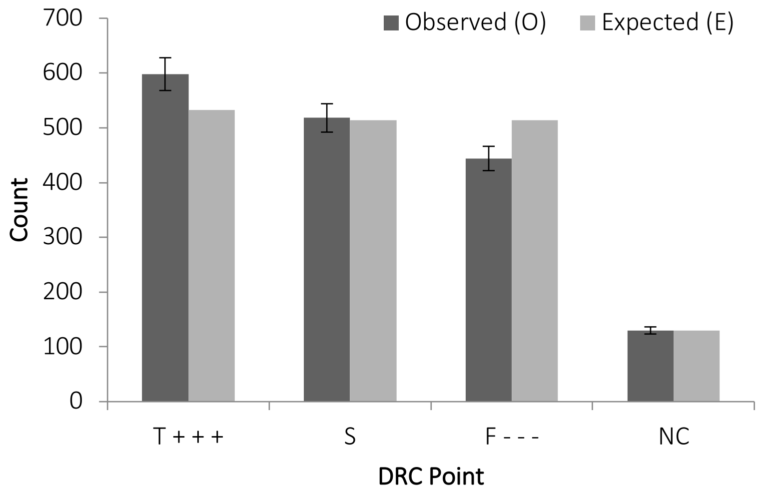


Figure 6.5: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for DRC point settings. (T: track; S: subgroup; F: full-sum, NC: no compression), --- indicates observed < expected a $p < .001$, +++ indicates observed > expected at $p < .001$. Error bars show ±5% error.

The frequency with which different DRC magnitude and type combinations were selected as preferred was calculated, pooling over the other independent variable (point), producing five settings: Moderate Compression (MC), Moderate Limiting (ML), Heavy Compression (HC), Heavy Limiting (HL), and No Compression (NC). Some settings were found to be selected at a frequency that differed significantly from chance [χ^2(4, N = 1690) = 23.97, $p < .001$], Fig. 6.6. Post-hoc pairwise comparisons show that HC was selected more frequently than expected ($p = .02$), whereas HL was selected less frequently than expected ($p < .001$). MC, ML and NC were not selected at a frequency that differed significantly from chance. These findings survive Bonferroni correction for two comparisons.

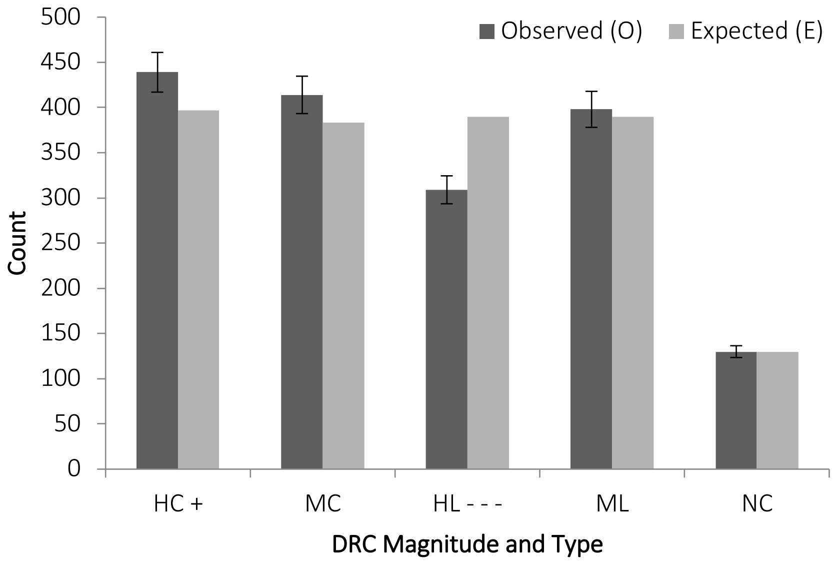


Figure 6.6: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for DRC magnitude and type settings combined. --- indicates observed < expected at $p < .001$, + indicates observed > expected at $p < .05$. Error bars show ±5% error.

The frequency with which different DRC type and point combinations were selected as preferred was calculated, pooling over the other independent variable (magnitude), producing seven settings: Compression Track (CT), Compression Subgroup (CS), Compression Full-sum (CF), Limiting Track (LT), Limiting Subgroup (LS), Limiting Full-sum (LF), and No Compression (NC). Some settings were found to be selected at a frequency that differed significantly from chance [$\chi^2(6, N = 1690) = 35.21$, $p < .001$], Fig. 6.7. Post-hoc pairwise comparisons show that CT was selected more frequently than expected ($p < .01$), whereas LF was selected less frequently than expected ($p < .001$). CF, CS, LT, LS and NC were not selected at a frequency that differed significantly from chance. These findings survive Bonferroni correction for two comparisons.

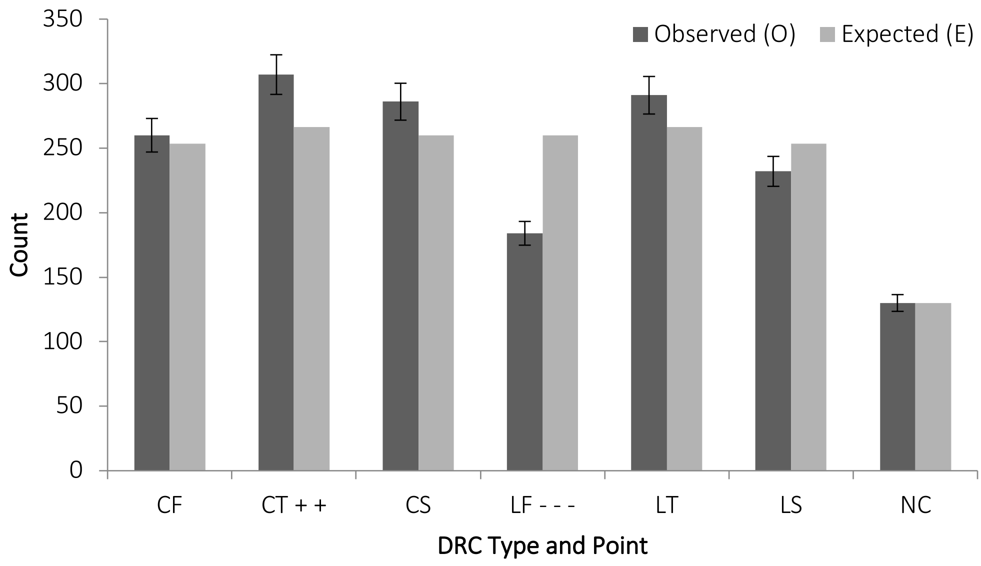


Figure 6.7: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for DRC type and point settings combined. (CF: compression & full-sum; CT: compression & track; CS: compression & subgroup; LF: limit & full-sum; LT: limit & track; LS: limit & subgroup; NC: no compression), --- indicates observed < expected at $p < .001$, ++ indicates observed > expected at $p < .01$. Error bars show ±5% error. Each observed count is located on the base of each corresponding bar.

The frequency with which different DRC magnitude and point combinations were selected as preferred was calculated, pooling over the other independent variable (type), producing seven settings: Moderate Track (MT), Moderate Subgroup (MS), Moderate Full-sum (MF), Heavy Track (HT), Heavy Subgroup (HS), Heavy Full-sum (HF), and No Compression (NC). Some settings were selected at a rate that differed significantly from chance [$\chi^2(6, N = 1690) = 38.46$, $p < .001$], Fig. 6.8. Post-hoc pairwise comparisons show that HT was selected more frequently than expected ($p < .01$), whereas HF was selected less frequently than expected ($p < .001$). HS, MF, MT, MS and NC were not selected at a frequency that differed significantly from chance. The findings survive Bonferroni correction for two comparisons.

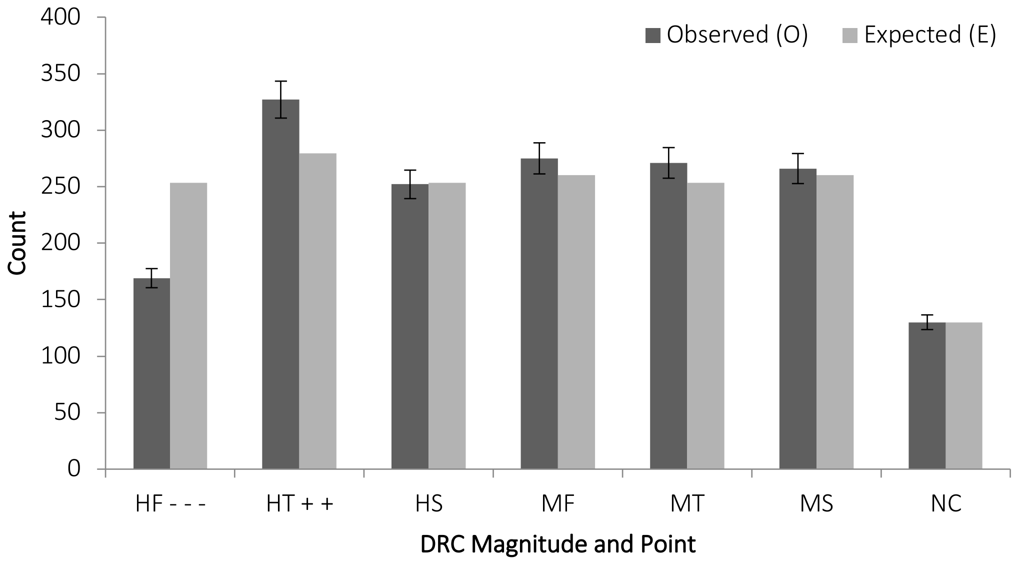


Figure 6.8: Listener preferences (observed [dark gray] *vs*. expected [light gray]) for DRC magnitude and point settings combined. (HF: heavy & full-sum; HT: heavy & track; HS: heavy & subgroup; MF: moderate & full-sum; MT: moderate & track; MS: moderate & subgroup; NC: no compression), --- indicates observed < expected at p < .001, ++ indicates observed > expected at p < .01. Error bars show ±5% error. Each observed count is located on the base of each corresponding bar.

Results are summarised in Fig. 6.9, in which significant and marginally significant stimuli/stimulus groups that were/were not preferred are shown.

	Preferred	**Not Preferred**
significant at p ≤ .05	C, CT, HCT, T, HT, HC	L, HL, HLF, F, LF, HF
marginal	HCS	HLS

Figure 6.9: Summarised listener preference results following post-hoc pairwise comparisons.

6.4 Discussion

Results indicate that the following stimulus groups were selected less often than expected by chance: L, HL, HLF, F, LF, and HF. A relationship between these stimuli is clearly apparent: many involve heavy magnitude limiting applied to the full-sum. This is as one might expect, given that it is this combination of settings that would be most likely to introduce distortion to the signal. One stimulus groups selected less often than chance that did not survive Bonferroni correction (HLS) had similar characteristics, entailing heavy magnitude limiting, but here applied to subgroups.

Results also indicate that the following stimulus groups were selected more often than expected by chance: C, CT, HCT, T, HT and HC. One stimulus did not survive Bonferroni correction (HCS). Again, these preferences are closely related, many entailing DRC applied to fewer signals (track or subgroup), and the use of compression rather than limiting. These findings are in agreement with some earlier studies in which compression was found to be preferred over no compression (Maddams, Finn and Reiss, 2012; Ma et al., 2015), although Ma et al. (2015) do stipulate that listeners may prefer no compression, contingent upon genre. These findings are consistent the hypothesis that DRC applied to fewer signals simultaneously (i.e., to tracks, rather than subgroups or the full-sum) produces the most agreeable results for listeners. Listener preferences for compression over limiting may be linked to the perceptual impact of nonlinear distortion: the higher DRC ratios used in limiting may increase the degree to which nonlinear frequency components are generated in the original signal, potentially causing it to sound increasingly “harsh” (Moore et al., 2004), and reduce “pleasantness” (Croghan, Arehart and Kates, 2012). A preference for light DRC over no DRC in the UNEQ condition of (Croghan, Arehart and Kates, 2012) was also reported. Considering that the function of DRC is to reduce variation in loudness, perhaps heavy DRC, in addition to introducing unpleasant distortion, also diminishes the expressiveness of music (Juslin and Laukka, 2003).

In Hjortkjær and Walther-Hansen (2014), no evidence of listener preference for “less compressed music” was found, in contrast to the finding reported here that compression was preferred over limiting (and no compression). It may be that moderate DRC is preferred over no DRC because the sonic characteristics imparted by DRC were pleasing to listeners for the musical stimuli used in this study. Potentially, this is a learned preference resulting from the widespread use of DRC in popular music production, making the sound of 'no DRC' high-fidelity audio sound less familiar and consequently less agreeable.

Listener preferences for DRC applied to individual tracks rather than the full-sum may be because DRC (whether limiting or compression) reduces audio fidelity to a lesser degree when fewer signals interact simultaneously by restricting the introduction of the aforementioned sum and difference components (Roads, 1979). The suggestion of Pestana and Reiss (2014) that DRC is justified for the compensation of "erratic loudness ranges" is reasonable, although the results reported here do appear to be at odds with their recommendation that DRC be used on the "overall mix" (referred to in this study as full-sum) as best practice.

In this study, DRC attack and release parameters were set to function in a generalised way, rather than being specifically optimised for each stimulus configuration, which may have influenced listener preferences. However, it would be difficult to obtain quantitative/comparable results had subjective adjustment of these parameters been undertaken, due to the variability of the individual signal envelopes. Furthermore, the organisation of instruments into subgroup was based on instrument type (e.g., drums, vocals, and other instruments). There are numerous ways in which the subgroups could have been organised which may have influenced DRC behaviour, and therefore listener preferences. However, the subgroups used in this study were broadly consistent with professional sound engineering preferences (for drum and vocal subgroups), reported in (Ronan et al., 2015; Ronan, Gunes and Reiss, 2017).

This experiment only tested listeners' preferences for different DRC configurations using one popular music excerpt (unknown to participants beforehand), restricting the degree to which we can generalise the findings reported here to other genres, or indeed to other stimuli within the popular music genre. However, using a single unknown musical excerpt (with different DRC configurations) as a source stimulus enabled participants to make a detailed audio quality comparison that would have been compromised had more audio stimuli (e.g., from multiple genres), or familiar stimuli, been used. Furthermore, the experiment required considerable concentration from participants, and the introduction of multiple musical clips would have rendered it too long to expect that concentration could be sustained (as would full counterbalancing, wherein every stimulus was compared to every other stimulus in both sequential orders by all participants). Informal feedback from participants suggested that the duration of the test protocol already may have been about as long as they were willing to tolerate. However, since the order that stimuli were used was partially counterbalanced (such that each participant was allocated one comparator stimulus), fatigue/boredom effects are not expected to have unduly influenced the pattern of results reported.

It is possible that the synchronicity of instruments may cause DRC to be invoked as a result of an ensemble of signals interacting constructively, rather than as a result of a single loud signal, which may lead to different pattern of results within genre, depending upon the timing and phase relationships between signals. Further work is required to verify whether the results reported here generalise to different music genres (i.e., to a wider range and combinations of timbres), durations, instrumentations, and musical structures, and to examine the impact of DRC ratio, attack, and release parameters on listener preferences more exhaustively.

6.5 Conclusions

By manipulating the point in the mix chain at which DRC was applied, this study supports the hypothesis that listeners prefer music with DRC applied to fewer signals simultaneously (i.e., to tracks prior to grouping/summation), which is expected to have reduced distortions associated with the application of DRC to pre-mixed signals (Roads, 1979; Zölzer, 2011; Moore et al., 2004; Croghan, Arehart and Kates, 2012; Wendl and Lee, 2014; Vickers, 2010; Hjortkjær and Walther-Hansen, 2014; Arehart, Kates and Anderson, 2011; Stone et al., 2009). The findings reported here also suggest that listeners prefer compression over limiting, and the use of moderate DRC over none. These findings are compatible with those of Croghan, Arehart and Kates (2012) who found that heavy DRC applied to maximise loudness reduced listener preference relative to where moderate DRC was used. In current industry practice, the application of compression to subgroups is commonplace; furthermore, limiting is often applied at the end of the signal chain, not just for overload protection, but also to increase loudness. Conversely, the findings reported here suggest that listeners prefer music to which DRC is applied early in the signal chain, and where compression (rather than limiting) is used.

In this next chapter, the potential impact DRC settings identified as being preferred and non-preferred on the results of a time estimation experiment is examined. This experiment aims to determine whether non-preferred DRC settings, as well as producing disagreeable distortion, might elicit fatigue effects.

Chapter 7:

The Effect of DRC on Listener Fatigue

7.1 Introduction

In the last chapter, an experiment was conducted to establish listener preferences for musical stimuli subject to 13 different DRC configurations, to determine the impact, in isolation and combination, of DRC magnitude, type and point settings. Significant differences in listener preferences were found, alluding to a dislike of heavy limiting applied to the full-sum signal, and other configuration variants that one would expect to be most deleterious to signal quality. In this chapter, an experiment is presented in which the estimated duration of a musical excerpt subject to each of the 13 different DRC configurations used in chapter five is measured in order to test the hypothesis that poorly controlled DRC can elicit listener fatigue or distraction, implied from a diminished ability to accurately estimate elapsed time, or an over/underestimation of the duration of stimuli with specific settings.

In this experiment, perceived elapsed time is measured while participants listen to music subject to different DRC configurations, enabling any change in perceived elapsed time as a result of particularly preferred or non-preferred (chapter five) DRC settings to be examined.

7.2 Method

7.2.1 Apparatus

The same apparatus as in chapter six was used, except that AKG K702 open-back dynamic reference headphones were used for playback, and were connected to an Apple Mac Pro Quad Core Tower running Pro Tools via a Metric Halo 2882 Audio Interface. An NTi Audio's Acoustilyzer AL1 A-weighted SPL meter was used to calibrate headphone output. Statistical analyses were performed using IBM SPSS.

7.2.2 Participants

The same participants recruited for the listening test described in chapter six were used again here (see section 6.2.2).

7.2.3 Stimuli

The same stimulus preparation routine described in chapter six was used (see section 6.2.3), except that a 20 s rather than 10 s musical excerpt was generated (Fig. 7.1). Like chapter six, stimuli subject to 13 different DRC configurations were prepared (see Table 6.1).

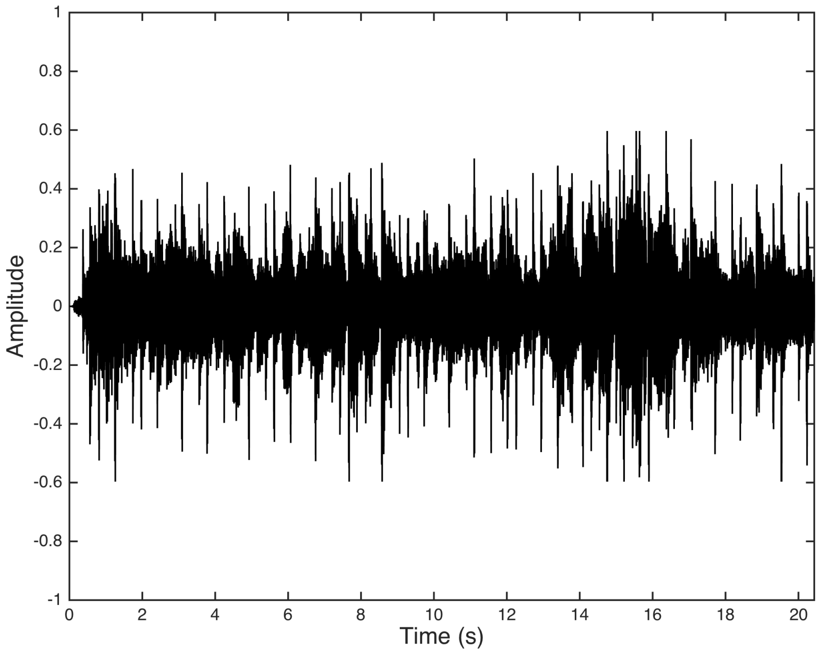


Figure 7.1: Time-domain plot of source stimulus (full-sum with no DRC applied).

7.2.4 Procedure

Participants sat in a quiet well-lit room, and their age and gender were recorded. They were informed that their task was to listen carefully to a musical excerpt, and that they would be asked to answer some questions afterwards. Next, the headphones and a blindfold were placed on their head. Headphone calibration was set to a maximum output of 78 dB (± 0.5 dB) SPL before data collection, using white noise generated in ProTools at -17 dB RMS, using the SPL Meter for final signal level adjustment. Once each participant had confirmed their readiness, stimulus playback was started. Once playback was complete, participants were asked the following four questions:

1. *Without the aid of a timepiece, please estimate the duration of the audio you just heard to the nearest second.*

2. *What is your evaluation of the sound quality of the audio you just heard, disregarding whether or not you like the piece?* [excellent | good | fair | poor | bad]

3. *Please indicate to what degree you agree with the following statement "The piece of music just played was enjoyable"* [strongly agree | agree | neutral | disagree | strongly disagree]

4. *Do you consider yourself to be an expert in music/audio quality (e.g., you work or study in this area)?* [yes | no]

The scale used in Question 2 was replicated from an EBU Technical Review described in Hoeg et al. (1997). Question 3 used a Likert-type scale (Likert, 1932). The entire experiment, including listening and responding to the above questions, took approximately five minutes per participant to complete. Importantly, the time estimation aspect of the experiment was not revealed to participants *a priori* to discourage counting and other possible explicit timing strategies.

7.2.5 Design

A between-subjects design was used, since each participant completed a single trial only. Since there were 13 stimuli, the recruitment of 130 participants ensured that there were an equal number of participants in each stimulus condition (i.e., 10 × 13). The allocation of participants to the 13 stimulus groups was random.

First, a set of one-sample t-test (two-tailed) were used to compare the time estimates provided by all participants (thus all stimuli) to the true duration of the stimulus. A two-tailed test was used to account for the possibility that estimates could be either longer or shorter than the true stimulus duration.

Next, estimates were grouped by DRC configuration (with 13 levels, see Table 6.1) to test the hypothesis that specific DRC configurations will elicit fatigue, implied by an impairment in the ability of participants to estimate elapsed time. However, as noted above, an alternative explanation for any difference in perceived elapsed time is that time appears to pass more quality when engaged in an enjoyable activity (Sackett et al., 2010). If this is the case, one would expect a correlation between reported enjoyment and the estimation duration of the musical stimulus.

Next, using the same general approach, time estimates were grouped by DRC magnitude (heavy, moderate and none), DRC type (compression, limiting, and none), and DRC point (channel, subgroup, full-sum, and none), collapsing across the other factors, to test the hypothesis that specific DRC settings (magnitude, type and point), independently of the other factors manipulated, alter the perception of elapsed time.

Next, pairs of factors grouped time estimates, similarly to Chapter 6 (magnitude and type; magnitude and point; type and point). Next, the above analyses were repeated after modifying the dependent variable from duration estimation to (absolute) duration estimation error from the true stimulus duration, in order that time estimation accuracy can be analysed independently of the polarity of the error.

Finally, the responses to the post data-collection questions were correlated with the duration estimation and duration estimation error dependent variables.

7.3 Results

A 25-bin histogram of the 130 duration estimations provides in this experiment is shown in Fig. 7.2, confirming that the data collected are approximately normal, justifying the use of parametric statistics. The mean of the 130 estimated stimulus duration responses collected was 23.84s (Sx 12.05).

Pooling across all 130 stimuli, a one-sample t-test comparing the estimated duration responses collected to the true duration showed a significant difference [$t_{(129)} = 3.63$, $p < .001$], with a mean overestimation of elapsed time of 3.84s.

Dividing the 130 time estimation responses into 13 groups of 10 responses (corresponding to each of the 13 stimuli) yields the data shown in Fig. 7.3. Performing a one-sample t-test to compare the estimated durations of each of the 13 stimuli relative to the true duration showed no significant differences for any single stimulus (all $p > .05$).

Sx.

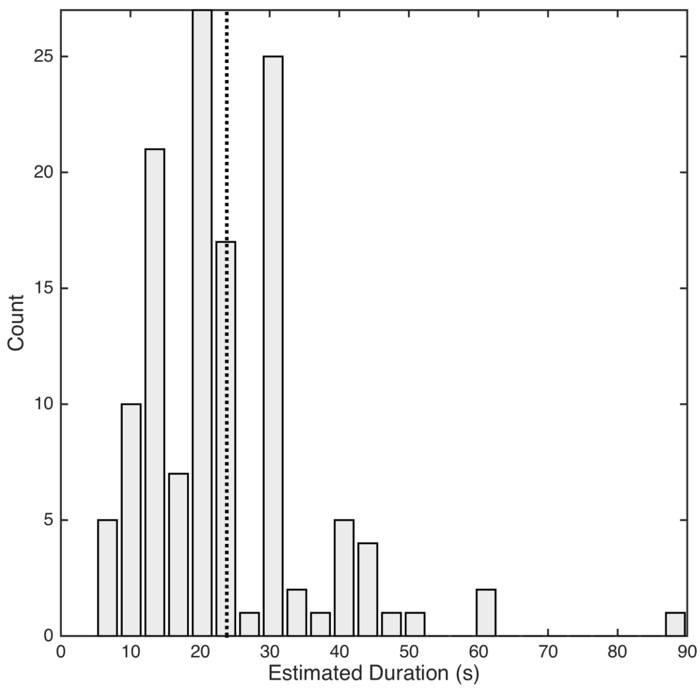


Figure 7.2: A 25-bin histogram of the 130 estimated duration responses collected. The dotted line denotes the mean estimated duration.

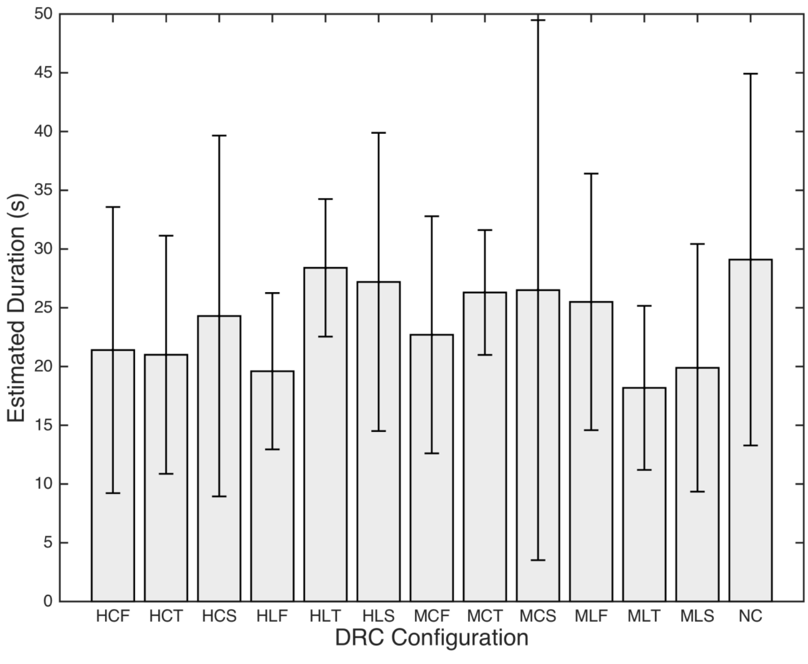


Figure 7.3: Mean estimated duration for each of the 13 stimuli. Errors bars are ±1 Sx.

Grouping stimuli according to DRC magnitude yields three groups: heavy (H), moderate (M) and no compression (NC). The results corresponding to this grouping scheme are shown in Fig. 7.4. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals significant differences for heavy [$t_{(59)}$ = 2.62, p = .01] and moderate [$t_{(59)}$ = 2.03, p = .05] DRC settings, but not for the no compression setting [$t_{(9)}$ = 1.82, p = .10].

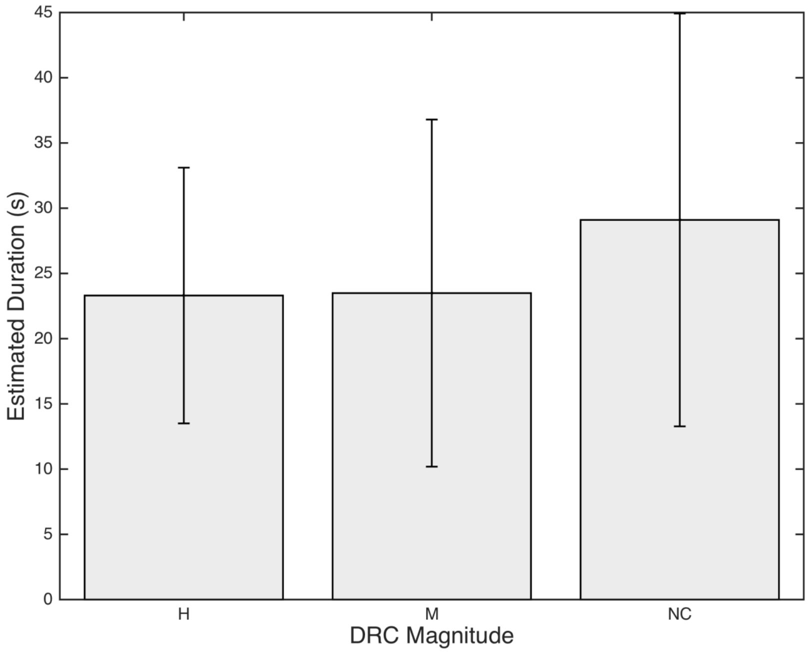


Figure 7.4: Mean estimated duration grouped by DRC magnitude. Errors bars are ±1 Sx.

Grouping stimuli according to DRC type yields three groups: compression (C), limiting (L) and no-compression (NC). The results corresponding to this grouping scheme are shown in Fig. 7.5. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals a significant differences for compression [$t_{(59)}$ = 2.92, p < .01], but not for limiting [$t_{(59)}$ = 1.44, p = .16], or the NC setting [$t_{(9)}$ = 1.82, p = .10].

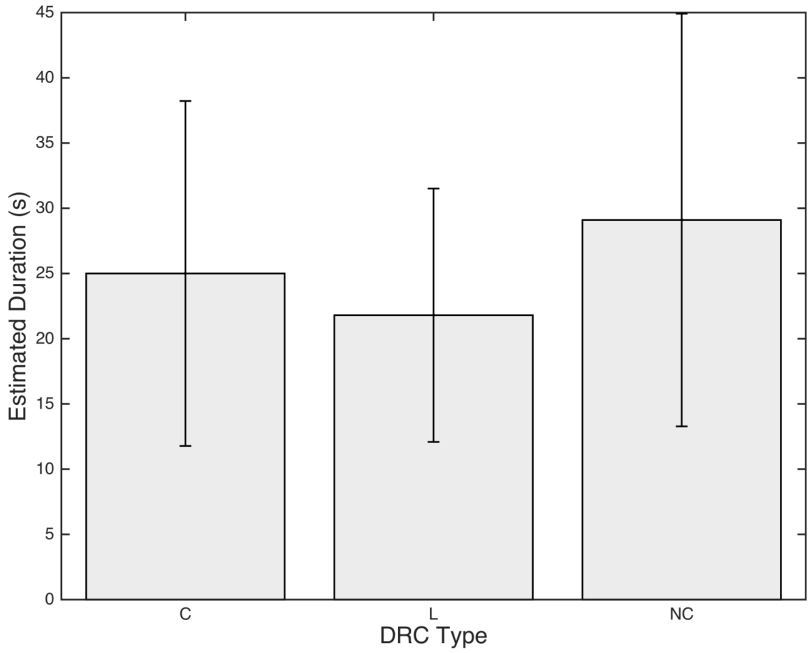


Figure 7.5: Mean estimated duration grouped by DRC type setting. Errors bars are ±1 Sx.

Grouping stimuli according to DRC point yields four groups: track (T), sub-group (S), full-sum (F), and no compression (NC). The results corresponding to this grouping scheme are shown in Fig. 7.6. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals a significant differences for sub-group [$t_{(39)}$ = 4.35, p < .001], but not for track [$t_{(39)}$ = 0.89, p = .38], full-sum [$t_{(39)}$ = 1.11, p = .27], or the NC setting [$t_{(9)}$ = 1.82, p = .10].

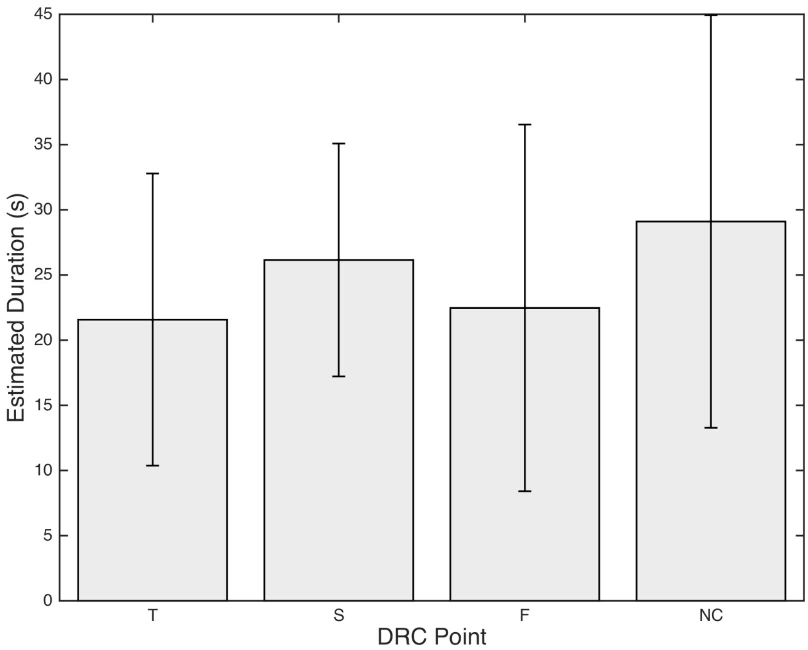


Figure 7.6: Mean estimated duration grouped by DRC point setting. Errors bars are ±1 Sx.

Next, the true duration of the stimulus was subtracted from each duration estimate, and the absolute taken, thereby quantifying time estimation error independently of the polarity of the error.

Dividing the 130 time estimation responses into 13 groups of 10 responses (corresponding to each of the 13 stimuli), using the dependent variable of duration estimation error, rather than raw estimated duration, yields the data shown in Fig. 7.7. Performing a one-sample t-test to compare the estimated durations of each of the 13 stimuli relative to the true duration showed significant differences for MCT [$t_{(9)} = 5.98$, $p < .001$], HCT [$t_{(9)} = 4.42$, $p < .01$], MLT [$t_{(9)} = 2.49$, $p = .03$], HLT [$t_{(9)} = 2.66$, $p < .03$], MCS [$t_{(9)} = 5.39$, $p < .01$], HCS [$t_{(9)} = 3.46$, $p < .01$], MLS [$t_{(9)} = 2.46$, $p = .04$], HLS [$t_{(9)} = 6.37$, $p = .0001$], HCF [$t_{(9)} = 3.16$, $p < .01$], MLF [$t_{(9)} = 3.89$, $p < .01$], HLF [$t_{(9)} = 2.78$, $p < .02$], and NC [$t_{(9)} = 3.89$, $p < .01$]. No significant difference was found for MCF [$t_{(9)} = 1.56$, $p = 0.15$].

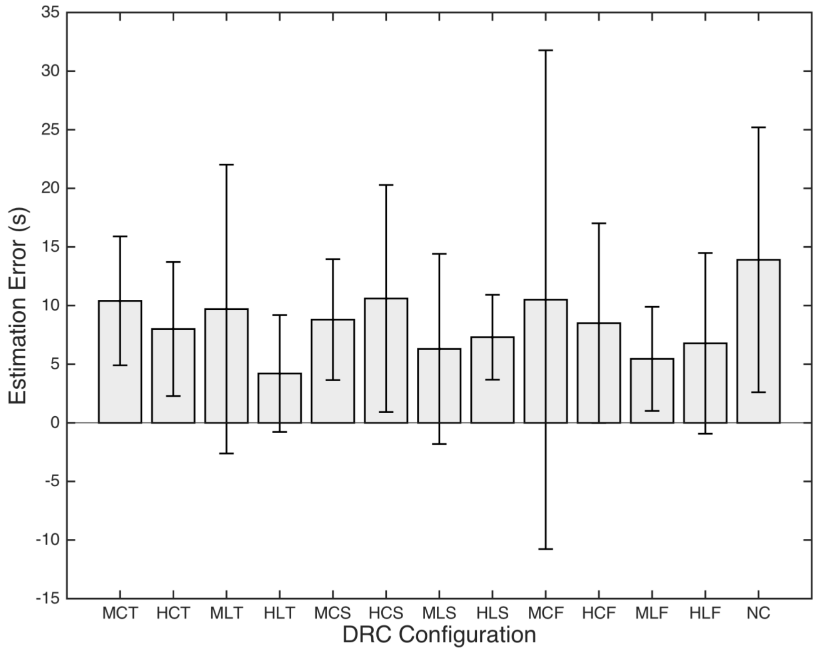


Figure 7.7: Mean estimation error for each of the 13 stimuli. Errors bars are ±1 Sx.

Grouping stimuli according to DRC magnitude yields three groups: heavy (H), moderate (M) and no compression (NC). The results corresponding to this grouping scheme, again using duration estimation error as the dependent variable, are shown in Fig. 7.8. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals significant differences for heavy [$t_{(59)}$ = 8.40, p < .0001], moderate [$t_{(59)}$ = 6.07, p < .0001], and NC [$t_{(9)}$ = 3.89, p < .01] DRC settings.

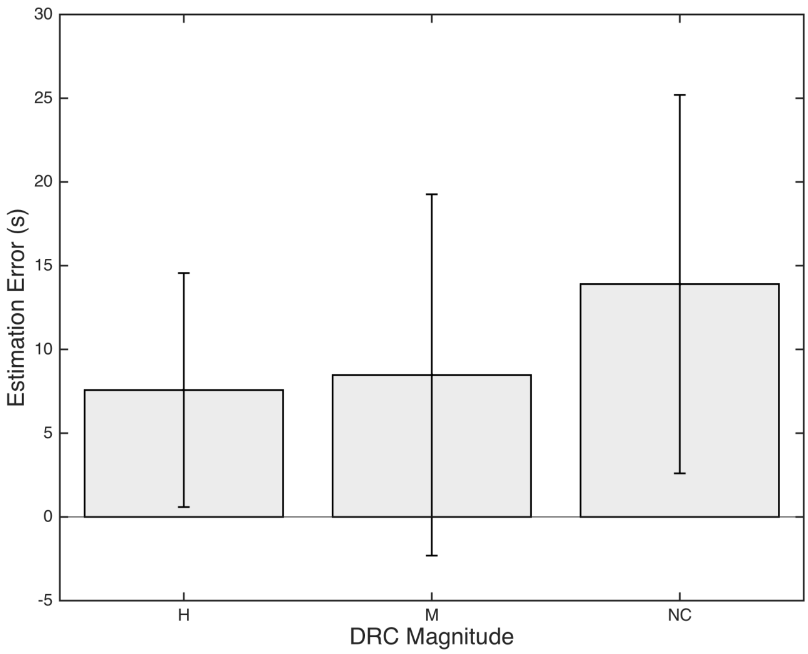


Figure 7.8: Mean estimated error grouped by DRC magnitude setting. Errors bars are ±1 Sx.

Grouping stimuli according to DRC type yields three groups: compression (C), limiting (L) and no compression (NC). The results corresponding to this grouping scheme, again using duration estimation error as the dependent variable, are shown in Fig. 7.9. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals a significant differences for compression [$t_{(59)} = 7.02$, $p < .0001$], limiting [$t_{(59)} = 7.06$, $p < .0001$], and NC [$t_{(9)} = 3.89$, $p < .01$] DRC settings.

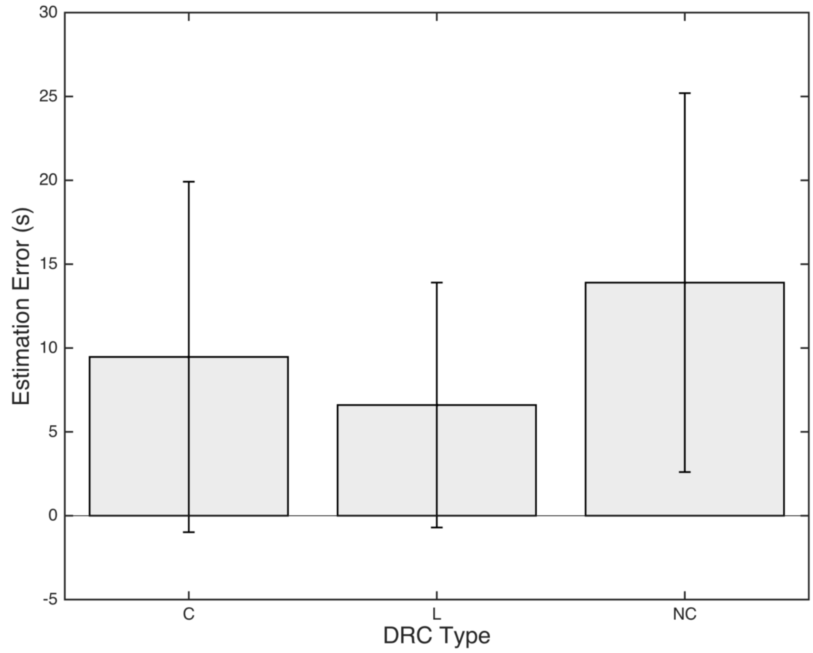


Figure 7.9: Mean estimation error grouped by DRC type setting. Errors bars are ±1 Sx.

Grouping stimuli according to DRC point yields four groups: track (T), sub-group (S), full-sum (F), and no compression (NC). The results corresponding to this grouping scheme, again using duration estimation error as the dependent variable, are shown in Fig. 7.10. A one sample t-test comparing the estimated duration responses in these groups to the true duration of the stimulus reveals a significant differences for track [$t_{(39)} = 6.53$, $p < .0001$], sub-group [$t_{(39)} = 7.48$, $p < .0001$], full-sum [$t_{(39)} = 4.12$, $p < .01$], and NC [$t_{(9)} = 3.89$, $p < .01$] DRC settings.

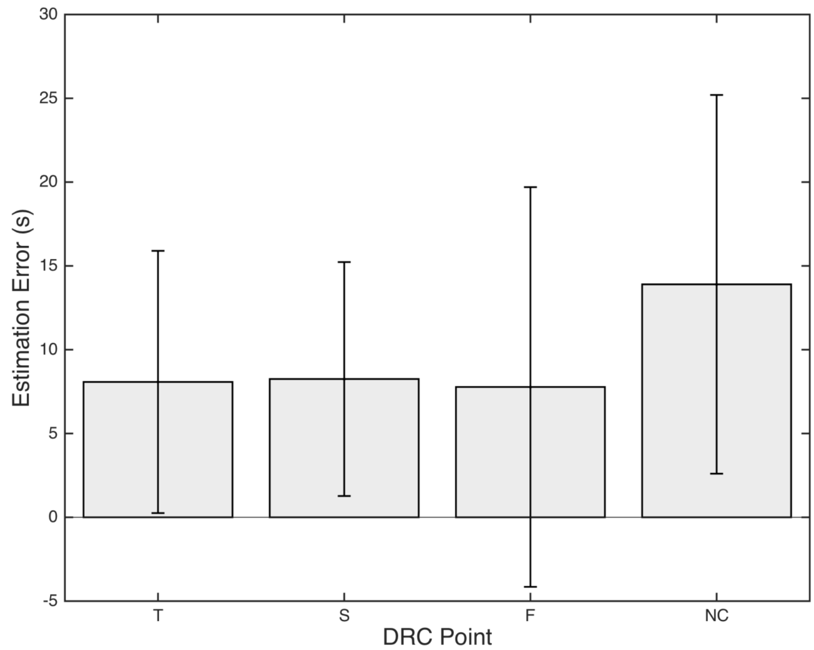


Figure 7.10: Mean estimation error grouped by DRC point setting. Errors bars are ±1 Sx.

Next, mean time estimation and mean absolute error were correlated with the responses to the non-binomial post data-collection questions. Scatter plots are shown in Fig. 7.11 and Fig. 7.12, for the 'enjoyment' and 'quality' judgments from the questionnaire. However, neither of the Pearson product moment correlation coefficients calculated were significant at $\alpha = .05$, using either the duration estimation [enjoyment $r_{(128)} = -0.06$, $p = .51$; quality $r_{(128)} = 0.04$, $p = .64$] or estimation error [enjoyment $r_{(128)} = -0.01$, $p = .89$; quality $r_{(128)} = 0.03$, $p = .76$] dependent variable.

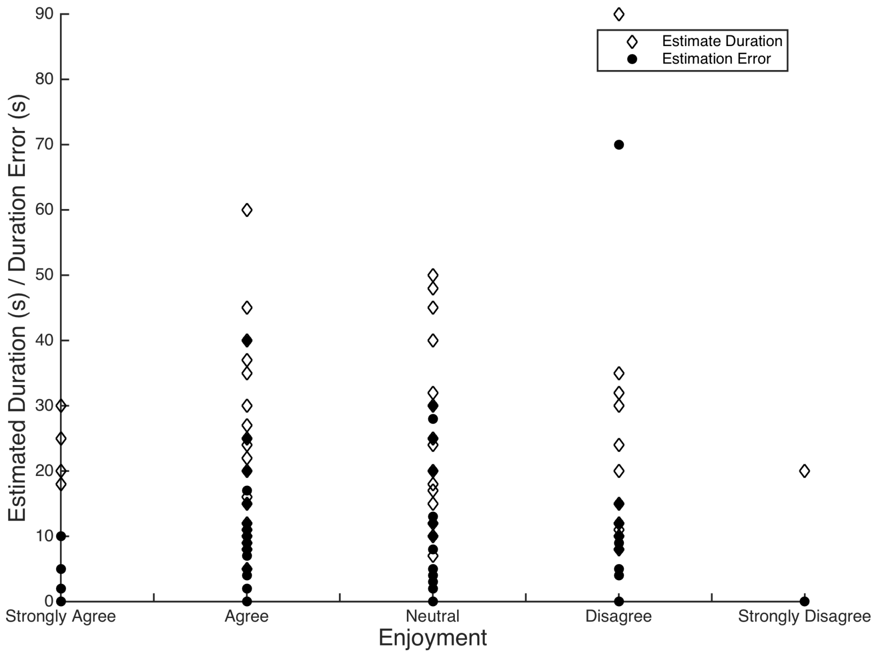


Figure 7.11: Scatter plot of raw estimated duration (hollow diamonds) and estimation error (filled circles) *vs*. enjoyment rating.

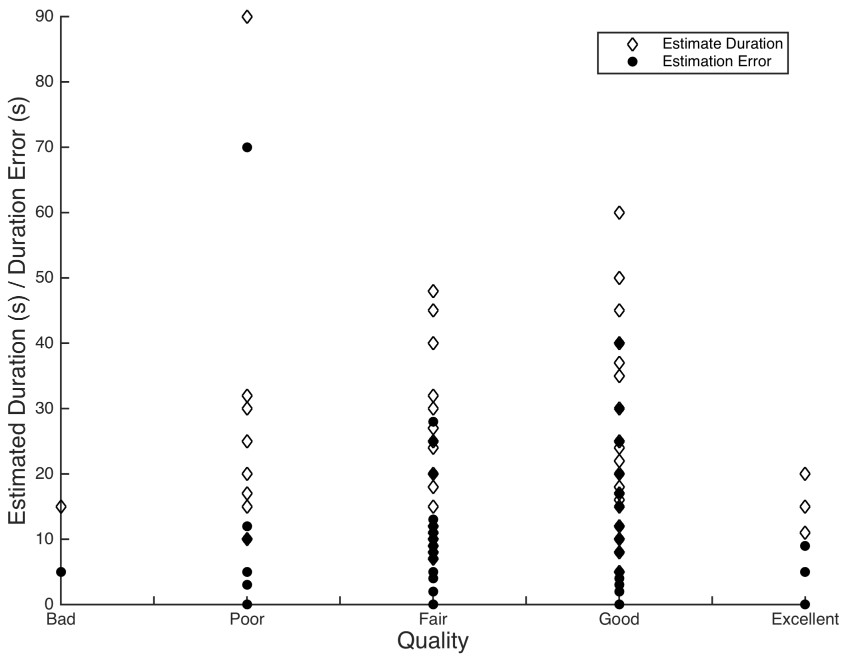


Figure 7.12: Scatter plot of raw estimated duration (hollow diamonds) and estimation error (filled circles) *vs*. quality rating.

No significant effect of listener gender was found for raw estimated duration [$r_{(128)}$ = -0.05, p = .55] or absolute error [$r_{(128)}$ = -0.05, p = .54] dependent variables. No significant effect of listener expertise was found for raw estimated duration [$r_{(128)}$ = -0.11, p = .21] or absolute error [$r_{(128)}$ = -0.16, p = .08] dependent variables. No significant effect of listener age was found for raw estimated duration [$r_{(128)}$ = 0.03, p = .75] or absolute error [$r_{(128)}$ = 0.04, p =.64] dependent variables.

7.4 Discussion

Pooled together all participants overestimated the duration of the stimulus (p < .001), but individually, no single stimulus yielded an average estimated duration that was still significantly different from the true stimulus duration (all p > .05). Grouping stimuli by DRC settings did produce time estimations that differed significantly from the true value.

Grouping by magnitude, both heavy (p = .01) and moderately (p = .05) compressed stimuli produced a significant overestimation of duration, whilst the no compression (p > .05) setting did not. However, more stimuli contributed to the heavy and moderate groups than the no compression group, so the lack of a significant effect for the latter group might potentially be the result of the smaller number of samples and consequently reduced degrees of freedom. Heavy stimuli were overall judged more accurately than the other two magnitude settings, though the difference with moderate is slight, contradictory with the findings of Stone et al. (2009) that DRC magnitude worsened participants' performance for some mental tasks.

Grouping by type, compression (p < .01) stimuli produced a significant overestimation of duration, whilst the limiting (p > .05) and NC (p > .05) setting did not.

Grouping by points, with sub-group (p < .001) stimuli producing a significant overestimation of duration, whilst the track (p > .05), full-sum (p > .05) and NC (p > .05) setting did not. Perhaps subgroup estimations were longer than full-sum and track stimuli because the subgroups configurations affected temporal judgment, there are numerous possible configurations (*see* De Man and Reiss, 2013; De Man et al. 2014; Ronan, et al., 2015; Ronan, Gunes and Reiss, 2017 for analysis of subgrouping practices).

Regarding duration estimation error, all tests apart from MCF ($p > .05$) produced significant absolute errors in duration estimation, including no compression. However, these results may be explained by the universally inaccurate estimations collected, rather than indicating any specific/interesting DRC-setting related trends.

An observation, perhaps unfounded, is that the results for magnitude and type setting appears to decrease as the amount of DRC increases. That is, heavy and limiting-type DRC parameters compress signals the most; they also had estimation means most similar to actual duration and estimated error, at the same time, moderate compression and NC (least of all in all analysis) had mean findings furthest away from the actual duration and estimated error. Perhaps the findings are more related to combinations of DRC. Stone et al. said that when applying DRC after signal summation (sum or more-so full) the combined effects of magnitude and point increased the negative effect on task performance.

Explanation for the universally inaccurate estimations was that the test duration was too short for participants to make realistic judgements. Brown (1997) utilised four continuous two-minute self-paced trials to minimise the effects of fatigue. North and Hargreaves (1999) test time was up to a maximum of 20 minutes. However, Zielinski, Rumsey, and Bech, (2008) advise that stimuli longer than 30 s may introduce bias to quality evaluations due to temporal and spectral variations over time, however stimuli durations are not discussed relative to time estimations tests.

This experimental design should have included single-task and dual-task scenarios in conjunction with temporal and non-temporal tasks similar to Brown. Particularly a design requiring participant-generated temporal reproductions, known to be associated with a reduction in the number of temporal cues perceived, therefore allowing a relatively longer time to pass before judging the specified interval has elapsed. However, this may prove difficult to incorporate with listening to DRC musical stimuli, unless the timing task was somehow aligned with the timing of the music.

One might expect the estimate duration of a musical excerpt to be affected by the DRC configuration used as a consequence of IMD, as other research has shown noise to negatively impact cognitive faculties (Rabbit, 1968; Sarampalis et al., 2009; Riley and McGregor, 2012). However, the SNR of the DRC stimuli may have been too high to induce fatigue-related symptoms. Rabbit (1968) utilised white noise correlated in intensity to voice amplitude. In Sarampalis et al. (2009) the noise level (four people speaking simultaneously in the background) was maintained at 65 dB SPL and the stimuli were adjusted to either -2 or 2 dB SNR. In Riley and McGregor (2012) white noise was mixed with the to-be-learned words at +8 SNR (recommended classroom SNR is +15 SNR (Flexer, 2002)).

The relatively low number of participants utilised in this experiment may have been a factor as there are only 10 participants in each stimulus group (although more in combined groups), meaning the analyses of individual stimuli may be underpowered. Rabbit (1968) tested a total of 293 participants over the three experiments (experiment 1: 80; experiment 2: 89; experiment 3: 124). Sarampalis et al. (2009) utilised a relatively small number of participants (experiment 1: 25; experiment 2: 25), but the coverage was extensive as each participant heard eight lists of 50 sentences (400 sentences in total for each participant). Brown (1997) utilised 94 participants for three experiments (experiment 1: 31; experiment 2: 33; experiment 3: 30); each experiment consisted two sessions (single-task and dual-task scenarios) of four continuous two-minute trials (16 minutes per participant), equating to 752 two-minute trials (25.13 hours of trials).

Rather than fatigue, the temporal judgements may have been the outcome of, distraction, enjoyment, or a combination of the three. Elapsed time can also be related to enjoyment – things seem to take longer when we're not enjoying them. For instance, Sackett, et al., (2010) concluded that *"… people often neglect the duration of events when judging* *hedonic value."* We might therefore expect poorer quality music to be less enjoyable, and to stretch time, relative to enjoyable music. The participants were only informed that they would be listening to a piece of music prior to the start of the experiment, and likely assumed they would be subjectively judging the stimuli. They were not informed they would be judging elapsed time. Therefore, perhaps the error bars are the real finding here that, as a whole, participants 'lost track of time'.

Perhaps there is an unseen relevance concerning the exhibited direction of shorter and longer verbal estimations. Droit-Volet et al. (2004) suggest that over-estimations may indicate emotional arousal (happy, sad, angry), while Danckert and Allman (2005) suggest underestimations may be indicative of boredom. Noulhiane et al. (2007) found negative sounds, and low-arousal stimuli judged longer than positive and high-arousal stimuli. Over and underestimations found in our test perhaps reflect the internal clock is adapting according to environment events (Droit-Volet and Gil, 2009) such as distortions associated with nonlinear processing.

NC (no-compression) perhaps challenges previous explanations, since it was found to deviate the furthest of all the stimuli from the true duration. NC lacks the nonlinear processing of the other stimuli, and therefore might be expected to exhibit an duration estimation closer to the true duration. Perhaps a partial explanation is that participants generally prefer compressed stimuli over NC (Maddams, Finn and Reiss, 2012; Ma et al., 2015). Kirk (1956) showed that untrained listeners who prefer lower-fidelity audio might do so because this matches their prior listening experiences; perhaps the NC stimulus produced a listening experience that participants were unaccustomed to, requiring more effortful listening, thereby yielding a greater duration estimation.

Although we compared NC to the other configurations, perhaps it was less vigorously tested as the other stimuli. Due to the nature of applying 'no-compression', i.e. having no compressor active, we viewed NC as 'full-mix with no compression' only. Although it makes no difference to the audio, potentially we should have also included NC as ‘channel mix no-compression’ and ‘subgroup mix no-compression’, therefore the statistical analysis three times more than it is. Note that NC was indeed balanced with all other comparisons, 10 tests for each DRC configuration (10 tests × 13 configurations = 130 total tests).

There were several factors tested (stimulus enjoyment, perceived stimulus quality, participants’ audio expertise, participants’ age group, and participants’ gender) which were thought to influence estimated duration according to the stimulus participants heard. Following is a discussion of each of these factors, though none had a significant impact on estimated duration or estimated error. In this experiment, neither stimulus enjoyment nor perceived stimulus quality had a significant effect on the resultant estimated duration or estimated error. The result is surprising, as one would assume that these subjective preferences would link to fatigue; as in unenjoyable, and poor quality, equate to fatigue. Perhaps subjects did not have time to form opinions on either enjoyability or quality in the 20 seconds of listening, though certainly, quality was noticeable. More likely though, these two questions were presented to participants' using the MUSRA (ITU-R BS.1534-1; multi-stimulus test with hidden reference and anchor) method of evaluating audio, known to introduce response mapping bias of implicit judgments. The process of mapping subjective experiences of audio to external responses is complicated and may comprise several strong biases (*cf*. Zielinski, Rumsey and Bech, 2008). The listening preference experiment that preceded this experiment utilised a two-interval forced choice stimulus evaluation method. Humans are naturally suited to comparative measurement. We are instinctively much better at looking at two things and comparing them than making an absolute judgment of stimuli one at a time. (Macmillan and Creelman, 1991; Lawless, 2013)

Kirk (1956) found that habituation likely effects participant’s perception of audio, and training can influence listening preferences either positively or negatively. These findings were supported by Olive (2011) and Olive (2012) who compared trained with untrained listeners, finding a reasonable relationship between expertise and general appreciation for audio fidelity.

They stipulated that untrained listeners possibly could not hear differences in audio fidelity or were simply less focused during the tests for lack of interested in the music presented. Reiss (2016) recommends training listeners so they can distinguish resolution differences in music. Although our experiment focused on the effect of DRC on time estimates, perhaps the results found here can be maximised by training listeners? Consequently, we found in this experiment, 'self-declared experts' in music/audio quality had no significant effect on the resultant estimated duration or estimated error and therefore no evidence of any correlation with cognitive effort.

Regarding age group, Olive's experiments tested high-school level (HS) (Olive, 2011) (mean age 16.5) against a trained panel (no demographics available other than participants' accumulated six months of experience in formal listening tests). Olive (2012) added three groups of college-level students (mean ages of: LMU, 20.1; UCI, 25.8; and CA, 25.3) to their previous experiment. They found significant differences between group preference choices, concluding the most experienced listeners (UCI) were the most accurate, followed by LMU, then HS, and the least accurate CA. They concluded that there was no statistical evidence showing teenagers and college students who regularly listen to low-bit-rate MP3s having a preference for low-quality audio. Again, very loosely relating poor audio quality to longer estimated duration, our results show no significant effect on the resultant Mean Time Estimation (MTE) according to stimuli, related to age group.

Finally, Barrett and Hodges (1995) found a significant difference between listening levels of male and female participants'. They found that middle school males preferred lower listening levels, and college level male and female students preferred similar listening levels. They also found that male music students in middle school and college habitually listened to music louder than their female counterparts, while college non-musical males habitually listened at lower levels than their non-musical female counterparts. Once again, if there were a tenuous link between preferred listening levels and estimated duration, this experiment shows no significant effect on the resultant estimated duration and stimuli relating to participant gender. Perhaps a good adjustment to our research in the future would be to allow participants to adjust the stimuli playback level to their preferred listening level, as Barrett and Hodges did, and then compare the resultant MTEs and listening levels.

7.5 Conclusions

This experiment tested participants' sense of perceived elapsed time relative to listening to test stimuli with various DRC configurations. The expected outcome was that DRC subjected to the most DRC would induce symptoms of fatigue evident by MTEs exceeding the actual 20-second musical stimulus length. The statistically significant findings suggest that all the test signals having DRC applied may be a source of fatigue. However, the results are inexplicable, demonstrating outcomes opposite to what one would intuitively expect and therefore are inconclusive at this time.

The experiment only tested one 20-second musical stimulus subjected to various DRC configurations. Future work includes extending this experiment to include different musical genres, lengthening the test and experimental design that includes single-task and dual-task scenarios in conjunction with temporal and non-temporal tasks like Brown (1997).

Chapter 8:

Analysis of the Effect of DRC on Simple and Musical Signals

8.1 Introduction

As discussed in section 2.3, some modern music productions have RMS amplitude levels (AES Standard) as high as –1.11 dBFS. Achieving this level of maximisation can be attained using heavy limiting with fast attack and recovery times (Izhaki, 2008; Katz, 2007). This chapter replicates the compression levels required to maximise the RMS of music programmes and increase loudness in a fashion comparable to modern popular music in order to understand the implications of heavy DRC usage.

In this chapter DRC is applied to simple signals and the nonlinear effects (*viz*. artefacts) introduced to these signal are examined. By observing the impact of DRC on simple signals, we may predict the effect they will have on more complex (e.g. musical) signals that are inherently more difficult to analyse. The test signals used include those in which the frequencies present are Harmonically Related (HRL) and Harmonically Unrelated (HUR), in order to evaluate the impact of DRC relative to musical tuning. Two methods of applying DRC are also compared. First, INDEP signals (independendently modulating; in which DRC is applied to channels prior to their summation), and second XMOD (cross modulating; in which DRC is applied to signals after summation), to gauge the impact of these different signal procesing chains on the magnitude of intermodulation distortion (IMD) introduced.

8.2 Method

The first test in this chapter is designed to test the effects of DRC on a single sine wave followed by combinations of sine waves modulating together.

8.2.1 Apparatus

Signals were generated using an Apple MacBook Pro running MATLAB. DRC was applied using the look-ahead brick wall limiter published in Zölzer (2011 pp.101-138). Signal analysis was performed using MATLAB, Sonic Visualiser (Cannam, Landone, and Sandler, 2010) using Libxtract: Loudness; RMS Amplitude (Bullock, and Conservatoire, U. C. E. B., 2007); and Spectral Centroid (Cannam, Landone, and Sandler, 2010).

8.2.2 Stimuli

Discrete (i.e., digital) sine wave stimuli of 1 second in duration, at frequencies of 110, 196 and 220 Hz were generated in MATLAB using formula Eq. 8.1, where f denotes frequency (in Hz), t is a vector of time indices (from 0 to 1), and x is a vector containing the corresponding sine amplitude values at each discrete index of t. Since the sine function was not multiplied by a scaling coefficient, the resultant amplitude range will be ±1.

$$x = \sin(2\pi f t) \qquad \textbf{Eq. 8.1}$$

The stimuli generated, had 16-bit resolution, and a sample rate of 44.1 kHz. HRL signals had frequencies 110 and 220 Hz and HUR signals had frequencies 110 and 196 Hz were generated (Fig. 8.1), using Eq. 8.2 in which f_1 and f_2 represent the two frequencies to be summed. In this study, the two sine waves were set to have equal amplitude (i.e., $a_1 = a_2$). Note that 220 Hz (A2) is an octave above 110 Hz (A1) in modern equal temperament standard concert pitch, and 196 Hz (G2) is ten semitones (a minor seventh) above A1.

$$x = a_1 \cdot \sin(2\pi f_1 t) + a_2 \cdot \sin(2\pi f_2 t) \qquad \textbf{Eq. 8.2}$$

The resultant signals were element-wise multiplied by a rising linear envelope, to trigger DRC, using Eq. 8.3 where x_1 is the original signal vector, r is the dimensionally-matched ramp vector, and x_2 is the resultant amplitude-ramped signal vector. The amplitude ramp started at 0 and terminated at ±1 at time 1 s.

$$x_2 = x_1 \circ r \qquad \textbf{Eq. 8.3}$$

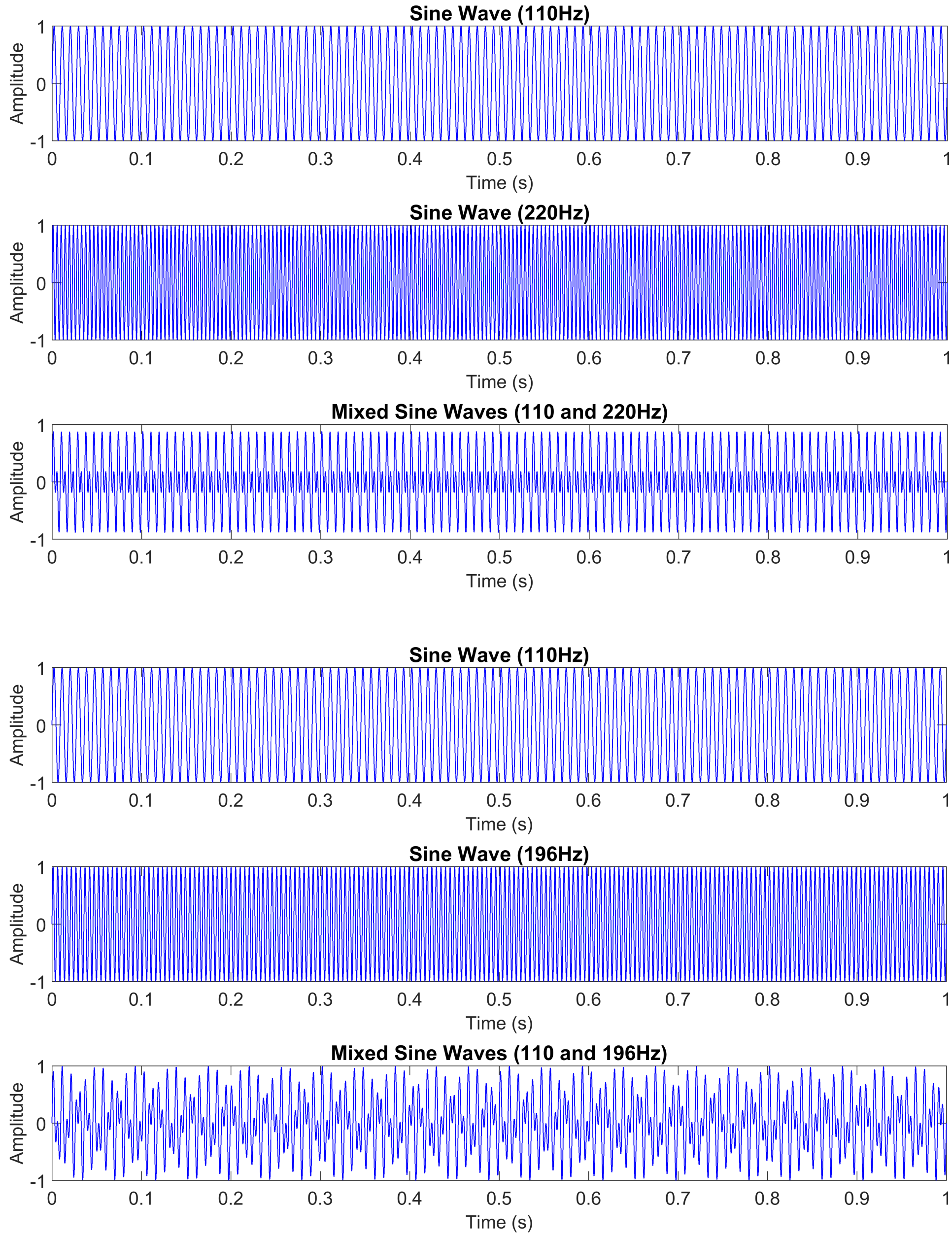


Figure 8.1: Sinusoidal test signals of 1 s duration, 16-bit resolution and 44.1 kHz sample rate, consisting of HRL 110 Hz and 220 Hz (top three panes) and HUR 110 Hz and 196 Hz (bottom three panes).

The DRC parameters used were: threshold 0.1, attack 0.3 ms, release 0.01 ms, look-ahead (delay) 5 ms, which produced a limiting effect. The delay was necessary to ensure that the limiter was fast-acting. The threshold multiplied by $20\log_{10}(|amplitude|)$ converts to standard logarithmic voltage units: 0.1 (-20 dB). Note that Test 2 utilises an alternative threshold of 0.3 to demonstrate the progressive nature of DRC.

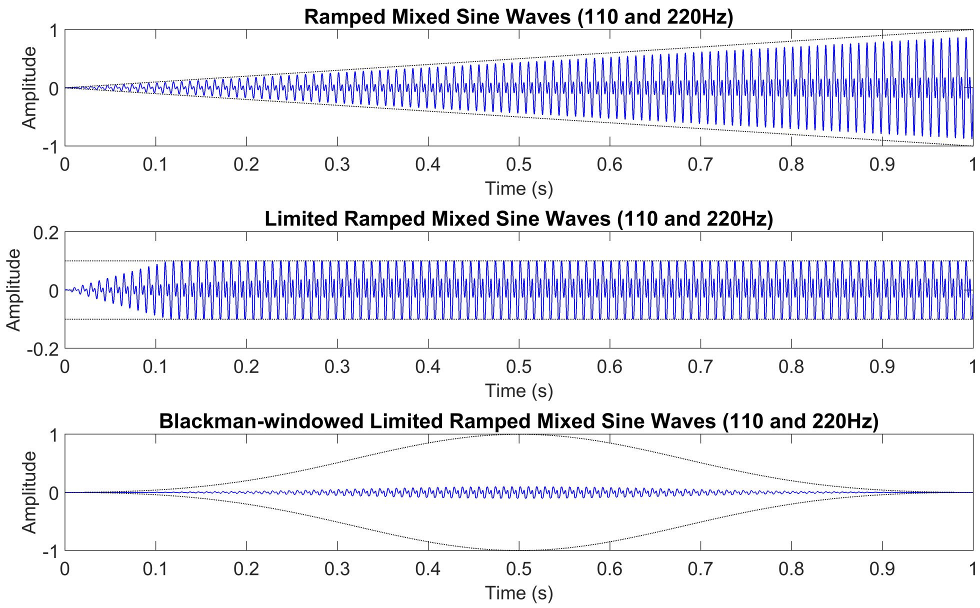


Figure 8.2: Summed sinusoidal test stimuli ramped (top pane), limited, and Blackman-windowed.

8.2.3 Procedure

In test 1 (Fig. 8.3), DRC (configured as described above) was applied to the single amplitude ramped 110 Hz sinusoidal signal, MATLAB plots of the resultant signal in both the time and frequency domains were generated (the latter using a Blackman windowed FFT), and spectral statistics were calculated as described below.

In test 2, DRC was applied to the amplitude ramped summed to XMOD sinusoidal signals (*cf.* Stone et al., 2009). Test 2A used HRL signals (Fig. 8.4); test 2B repeated the procedure utilising HUR signals (Fig. 8.5).

In test 3, DRC was applied to the amplitude ramped INDEP signals (*cf.* Stone et al., 2009) before summation. Test 3A used HRL signal (Fig. 8.6); test 3B used the HUR signals (Fig. 8.7).

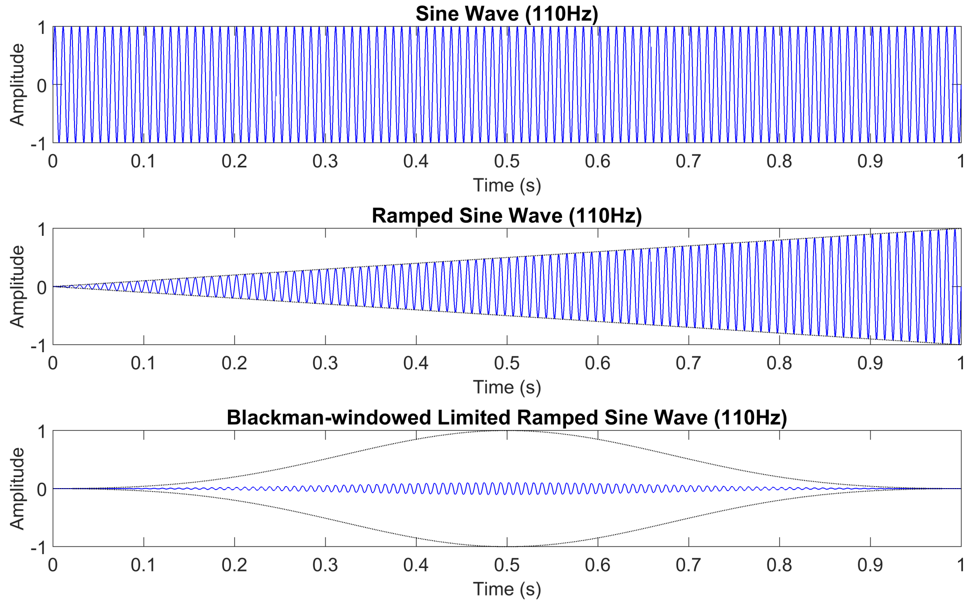


Figure 8.3: Test 1: 110 Hz sinusoidal test stimuli ramped (middle pane), limited and Blackman-windowed.

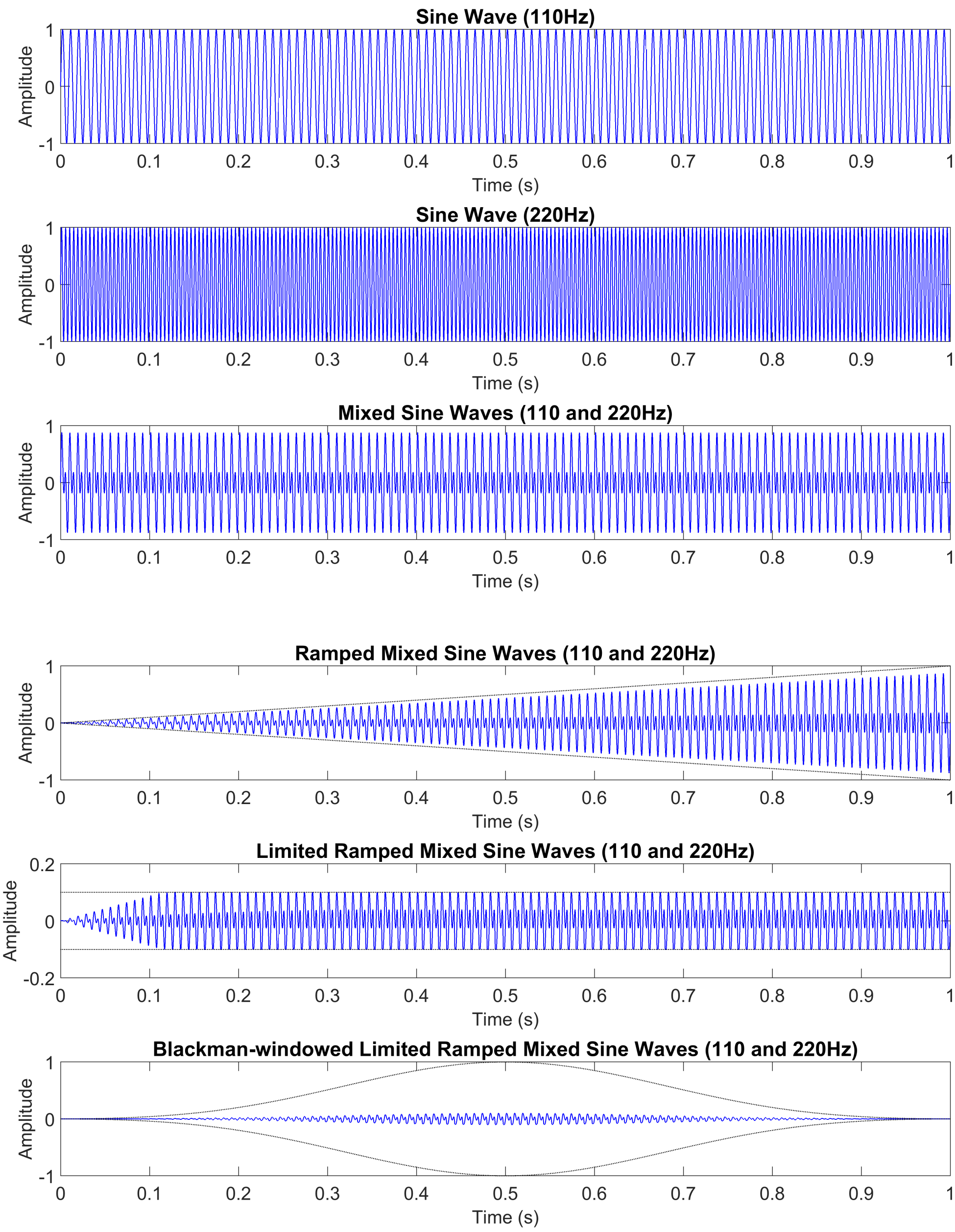


Figure 8.4: Test 2A: Time domain plots of the XMOD HRL signals (110 Hz and 220 Hz sine waves) prior to and post mixing. The test stimuli are ramped, limited and Blackman-windowed.

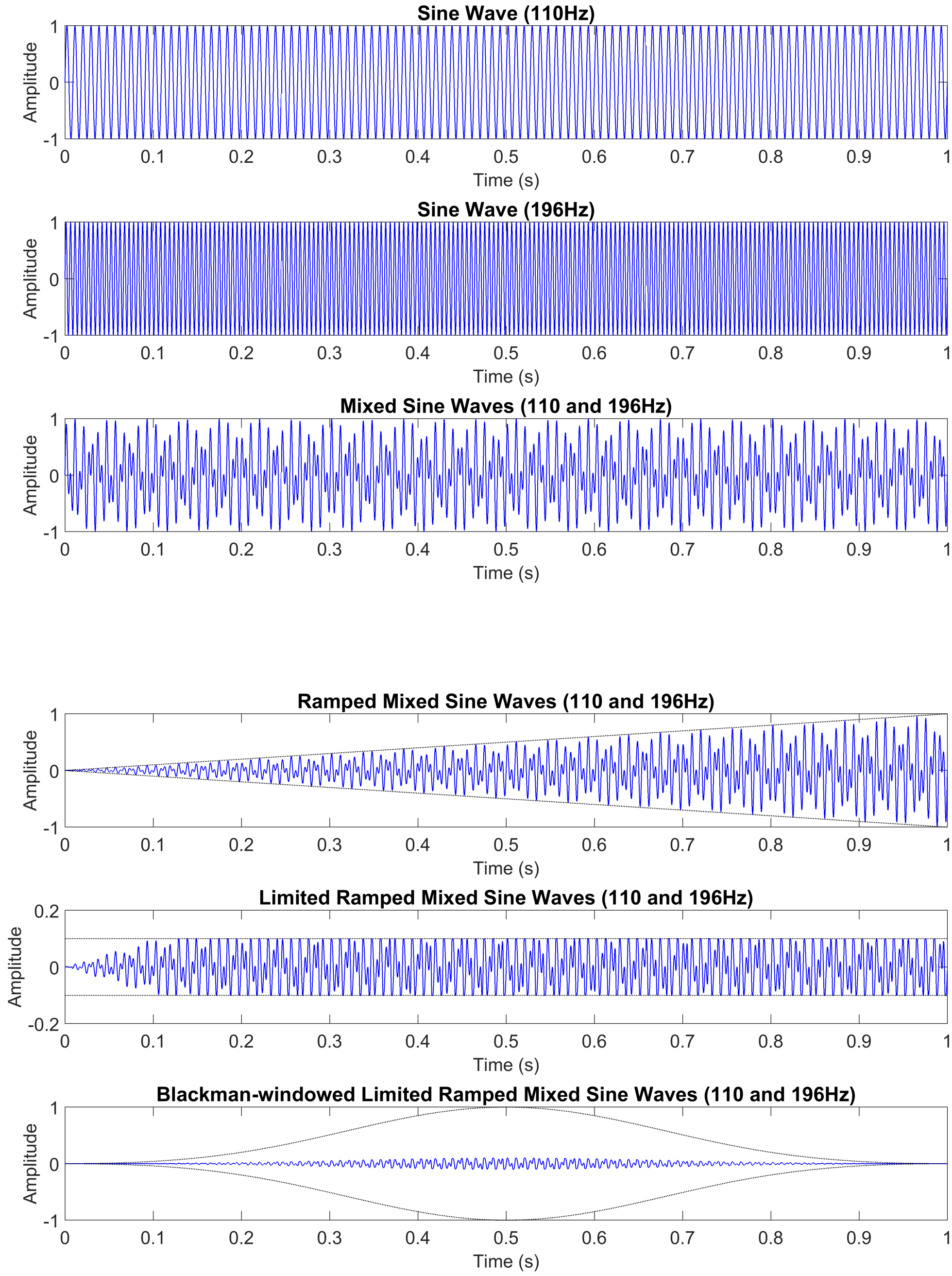


Figure 8.5: Test 2B - time-domain plots of the XMOD HUR signals (110 Hz and 196 Hz sine waves) prior to and post mixing (top three panes). The test stimuli are ramped, limited and Blackman-windowed (bottom three panes).

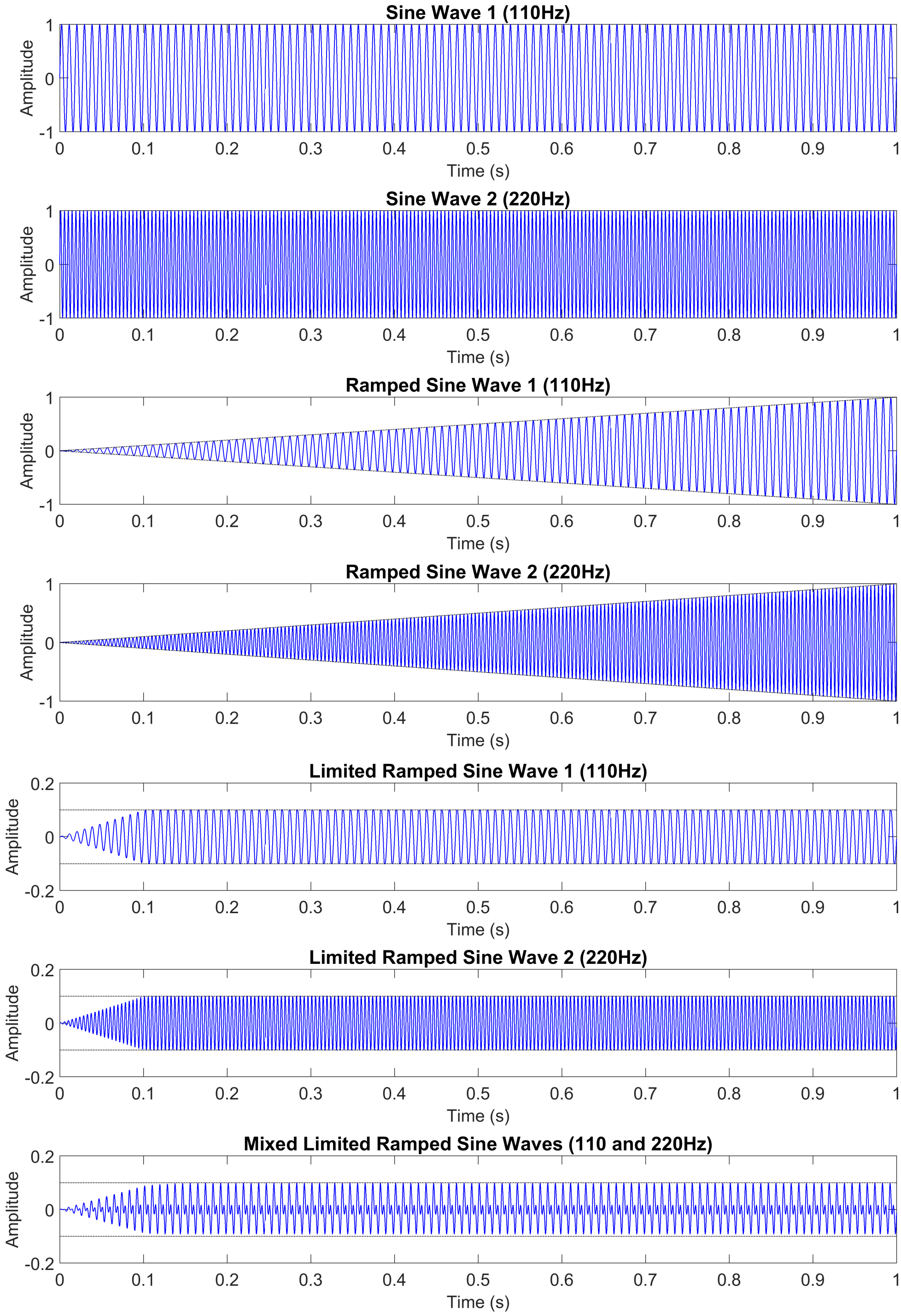


Figure 8.6: Time domain plots of the discrete INDEP HRL 110 Hz and 220 Hz signals (top two panes). The test stimuli are ramped (middle panes), limited and Blackman-windowed (bottom three panes).

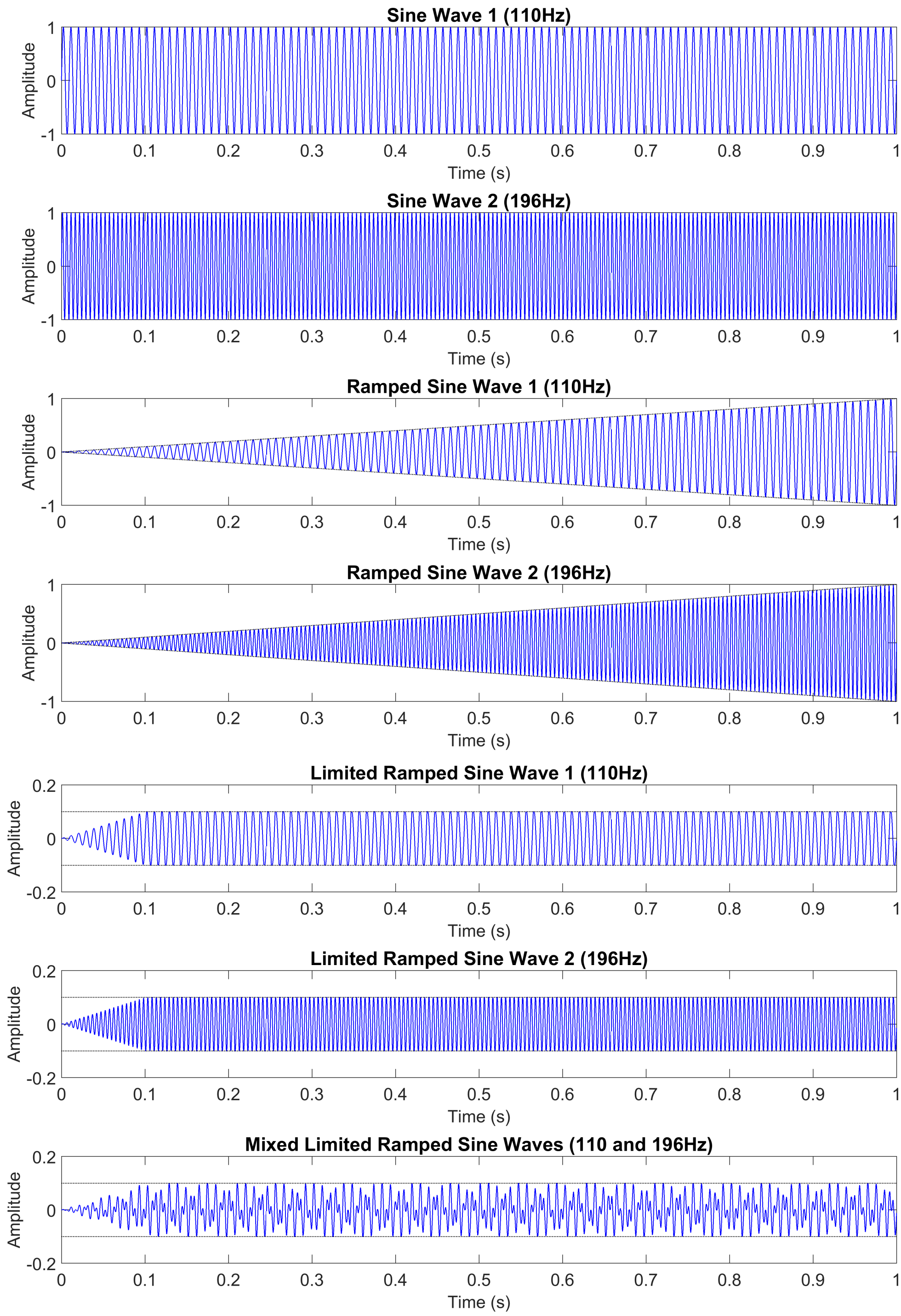


Figure 8.7: Time domain plots of the INDEP HUR 110 Hz and 196 Hz signals (top two panes). The test stimuli are ramped (middle panes), limited and Blackman-windowed (bottom three panes).

Further analysis of the resultant signals using Sonic Visualiser was performed, which produced additional spectrogram plots and signal statistics. Two other transform analyses tools employed were Bark Scale Loudness (BSLoud) and RMS amplitude.

The libXtract loudness (Bullock, 2008) transform design, which is based on the Bark Scale Loudness (BSLoud), a psychoacoustically-informed frequency scale, which ranges from 1 to 24 Barks corresponding to the first 24 critical bands of hearing (Zwicker, 1961). The model and is based on Peeters' (2004) adaptation of total and specific loudness as initially defined in of Moore, Glasberg and Baer (1997), returning a number 0 to 12 (12 being the maximum BSLoud) (*cf.* Bullock, 2008).

BSLoud is the loudness of the zth Bark band (Peeters, 2004). The total loudness is the sum of the individual loudness coefficients. RMS amplitude (Eq. 5.1), a statistical measure of dispersion that is an objective measure of (non-psychoacoustic) loudness, was used with a 0.01-second window in Sonic Visualiser. Windowed RMS provides information about a signal's power over time (Peeters et al., 2011).

Log Frequency Centroid (LFC) (Eq. 8.4) is a spectral centroid measure which provides the midpoint of spectral energy for each input frame. The usage here is as an indicator of distortion (i.e., higher frequency content added because of the addition of harmonic and inharmonic distortion introduced by DRC). Spectral centroid describes the spectral 'centre of gravity' of audio signals (Bullock, 2008). The spectral centroid is a measure of power distribution as a function of frequency (Schubert and Wolfe, 2006):

$$f_c = \sum_{k=1}^{n} f_k a_k \Big/ \sum_{k=1}^{n} a_k \qquad \textbf{Eq. 8.4}$$

Where a_k = spectral magnitude

f_k = frequency

f_c = the centre frequency of the bin

n = sample size

LFC (Eq. 8.5), a member of the spectral centroid family returning the centroid of the log-weighted frequency spectrum (Schubert, 2004), was used for identifying distortion in these experiments.

$$l_c = \log_{10}\left(\sum_{k=1}^{n} f_k a_k \Big/ \sum_{k=1}^{n} a_k\right)$$ **Eq. 8.5**

Where a_k = spectral magnitude

f_k = frequency

l_c = the log-weighted frequency centre frequency of the bin

n = sample size

Data collected from the three transforms was exported as text and entered into an Excel spreadsheet for analysis using Mean Absolute Error (MAE) and standard deviation (Sx) mean.

Mean absolute error (Eq. 8.6) measures how widely dispersed an array of numbers are from the mean, or it is a measure of accuracy between two arrays (0 to ∞); the smaller the error the better the model:

$$MAE = \frac{1}{n}\sum_{k=1}^{n} |y_k - x_k|$$ **Eq. 8.6**

Where y_k and x_k = sample means

n = sample size

Standard deviation (Sx) (Eq. 8.7) measures how widely dispersed an array of numbers are from the mean:

$$Sx = \sqrt{(x - \bar{x})^2}/(n - 1) \qquad \textbf{Eq. 8.7}$$

Where $\bar{x}$ = sample mean

n = sample size

8.3 Results and Discussion

8.3.1 Test 1: Single sine wave under DRC

The results of test 1 Blackman-windowing (Fig. 8.8) shows the frequency domain plot of a single fundamental frequency centred at 110Hz, corresponding to the frequency of the sine wave in the time domain. Application of Blackman-windowing did not add appreciable distortion as it is a linear process; which also holds true in the following tests using mixed signals. Therefore, hereafter Blackman-windowing linearity is assumed.

Test 1 Blackman-windowed amplitude ramped signal envelope (Fig. 8.9) frequency domain plot shows a single fundamental frequency centred at 110 Hz corresponding to the frequency of the sine wave in the time domain. Again, application of amplitude ramping did not add appreciable distortion as it is a linear process; which also holds true in the following tests using mixed signals. Therefore, hereafter amplitude ramping linearity is assumed.

Applying DRC (Fig. 8.10), a nonlinear process, adds distortion, manifesting as additional frequency components added to the original signal. The waveshaping function of DRC has the effect of ‘squaring off’ the compression and rarefactions of a sine wave (Fig. 8.11) and therefore resultant signal contains only odd integer harmonic components related to the fundamental frequency (i.e., at 330 Hz, 550 Hz, 770 Hz, and so on). These findings are as expected and consistent with the functionality of nonlinear processors as described in section 3.4 (Metzler, 2005; Roads, 1979; Zölzer, 2011; Moore et al., 2004).

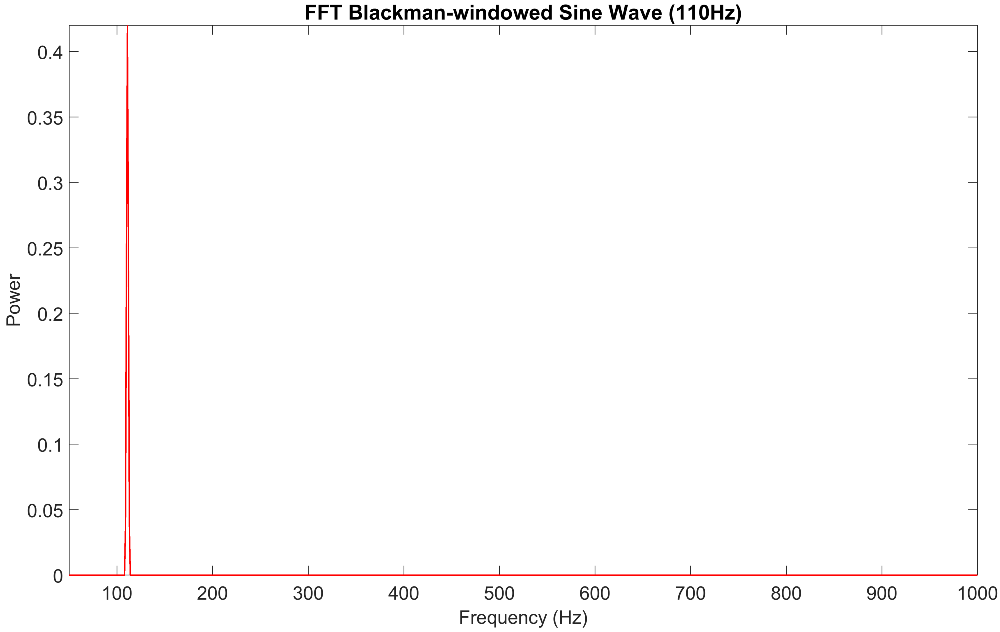


Figure 8.8: FFT results for Test 1 demonstrating singular frequency components (110 Hz) in the linear application of Blackman-windowing.

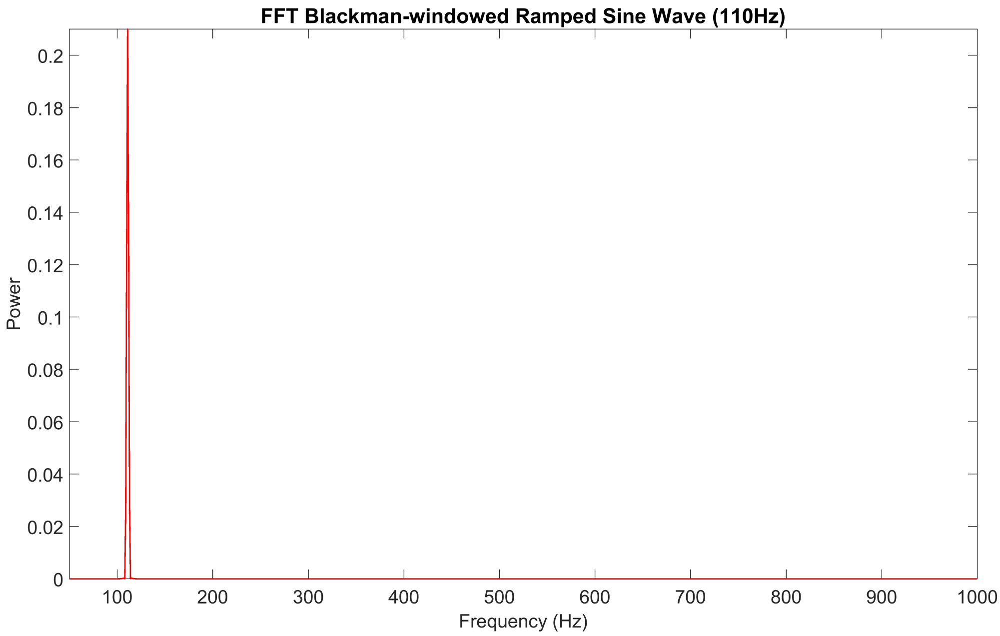


Figure 8.9: FFT results for Test 1 demonstrating singular frequency components (110 Hz) in the linear application of amplitude ramping.

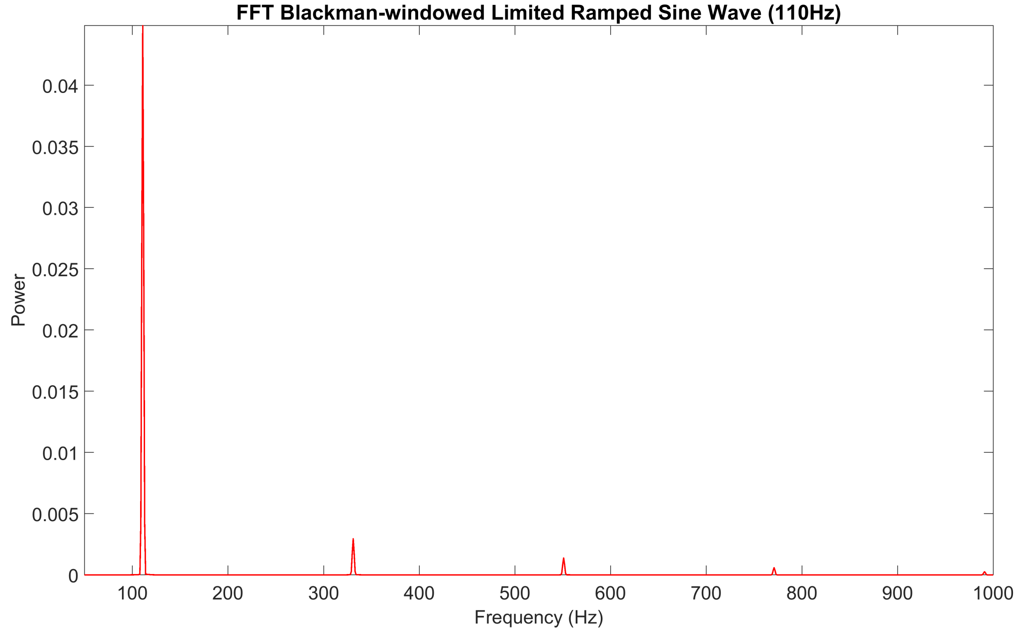


Figure 8.10: FFT results for Test 1 demonstrating signal distortion produced by the application of DRC to the 110 Hz sine wave.

Figure 8.11: Time-domain plots (top two panes) showing the ramped sine wave in the top pane and the experimental signal in the pane below (amplitude (dB) vs. time (ms)). Frequency-domain plots (bottom two panes) displaying frequency (Hz) and amplitude (dBFS) showing the ramped sine wave in the top pane and the experimental signal in the pane below.

Table 8.1 shows the result of statistical comparisons of the HL and NC signals. There is an absolute difference (ABS diff.) between the mean amplitude of the two signals of 19.37 dB, which is to be expected as DRC works by reducing the amplitude of the loudest portions of a signal only, unless being utilised as an expander. Another influencing factor is likely the ramped amplitude of the control signal whereas the test signal becomes more steady state once the DRC threshold limits the amplitude.

The MAE of 0.25 shows that applying DRC to a single sine does not alter the signal significantly.

Single Sine Wave	NC	HL
n (samples)	44101	44101
Mean (dB)	-56.79	-76.17
ABS diff. (dB)	19.37	
MAE	0.25	

Table 8.1: Statistical comparison of the 110 Hz sine wave with and without DRC.

Table 8.2 compares loudness, RMS amplitude, and LFC for the HL and NC stimuli. What is particularly interesting from the comparisons is that although the RMS amplitude between the two signals differs by about 14 dB, the loudness differential is only about 1 dB. Of course, the effect on loudness will be different for higher frequency test signals as normal hearing sensitivity to loudness increases at higher frequency bandwidths.

The mean of the measured LFC for the 110 Hz NC stimulus is accurate to about 1 Hz. The LFC measure demonstrates that the distortion components introduced by DRC introduced a frequency shift of 21.09 Hz to the HL test signal, shifting the LFC to 131.95 Hz. This shift in LFC is evident from the onset of DRC at 0.01 ms, as shown in Fig. 8.12. The shift is a consequence of the amplitude reduction (-14 dBFS) of the fundamental frequency and addition of distortion components. Equalising the RMS of the signals by applying 14 dB of gain to the HL signal would proportionally shift the loudness of the HL signal.

The MAE for loudness and RMS is low, indicating that applying DRC to a single sine does not alter these measures notably. However, the MAE for LFC (21.09) appears significantly altered, reinforcing the observations described above.

Single Sine Wave	**BSLoud**		**RMS Amplitude**		**LFC**	
	NC	***HL***	***NC***	***HL***	***NC***	***HL***
n (samples)	44.00	44.00	87.00	44.00	88.00	88.00
Mean	10.31 dB	11.26 dB	-9.05 dB	-23.04 dB	110.87Hz	131.95Hz
ABS Diff.	0.95 dB		13.99 dB		21.09 Hz	
MAE	0.48		.032		21.09	

Table 8.2: Statistical quantification of the properties of the 110 Hz sine wave with (HL) and without DRC (NC).

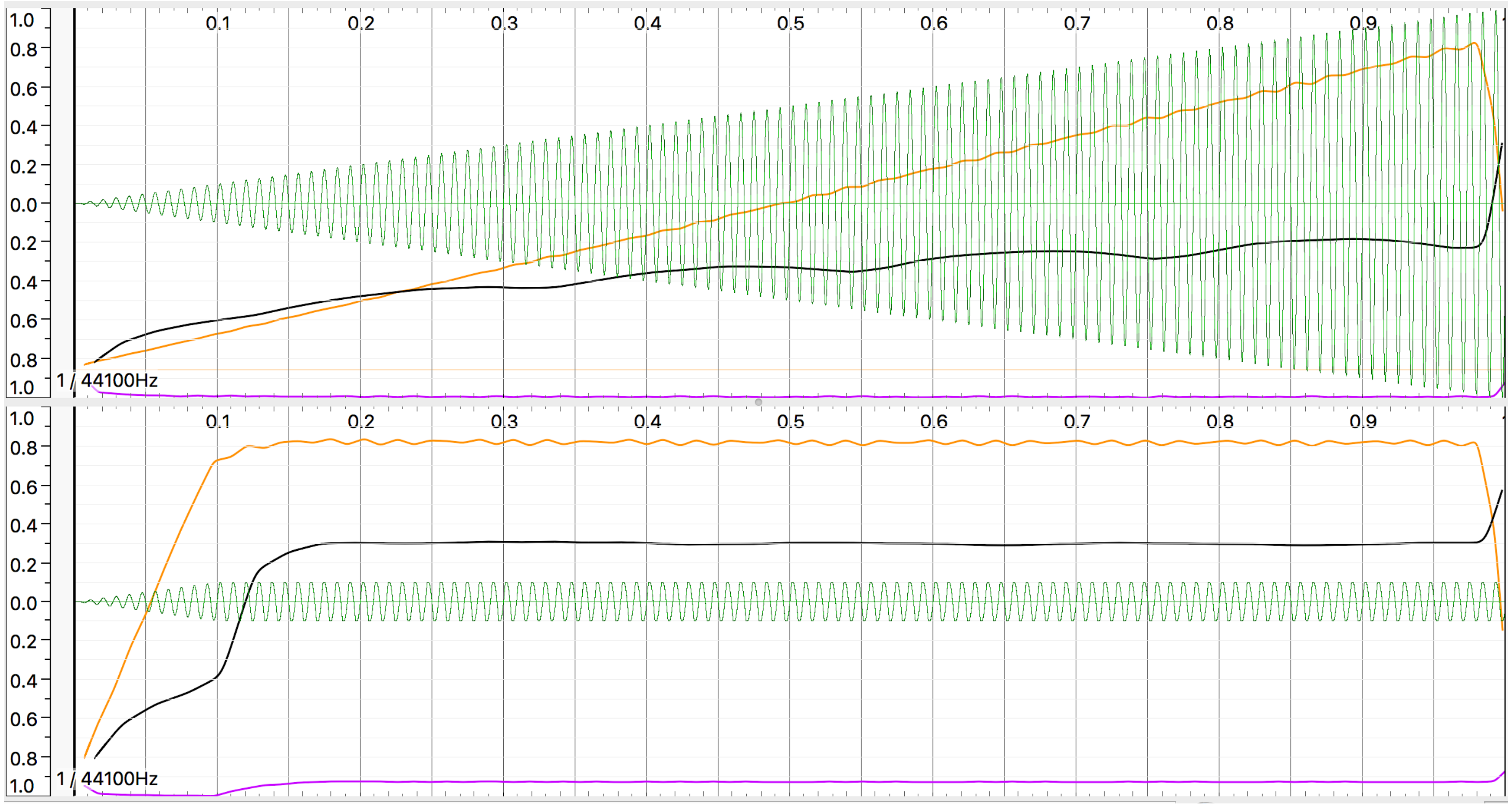


Figure 8.12: Amplitude vs. time (ms) plot showing the 110 Hz NC in the top pane and the 110 Hz HL in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

8.3.2 Test 2: DRC applied to cross-modulating (XMOD) signals

Test 2 has two parts: 2A applies DRC to XMOD signals constructed from HRL frequencies (110 Hz and 220 Hz) and 2B applies DRC to XMOD signals constructed from HUR frequencies (110 Hz and 196 Hz).

Test 2A used light (threshold = 0.3), and heavy (default threshold = 0.1) DRC magnitudes to illustrate the progressive nature of DRC. Visual inspection reveals similar results to test 1 (Fig. 8.13 and 8.14). Comparing the results of the two tests demonstrates that progressively increasing DRC magnitude produces progressively increased distortion components that are harmonic related to both fundamentals. Distortion components for the HRL signal are a spectrum of components whose frequencies are simple integer ratios (harmonics) of the fundamental frequency of each modulating signal (i.e., at 330 Hz, 440 Hz, 550 Hz, 660 Hz, 770 Hz, and so on). The resultant power units and peak amplitude decrease with increased DRC magnitude (lowering the threshold).

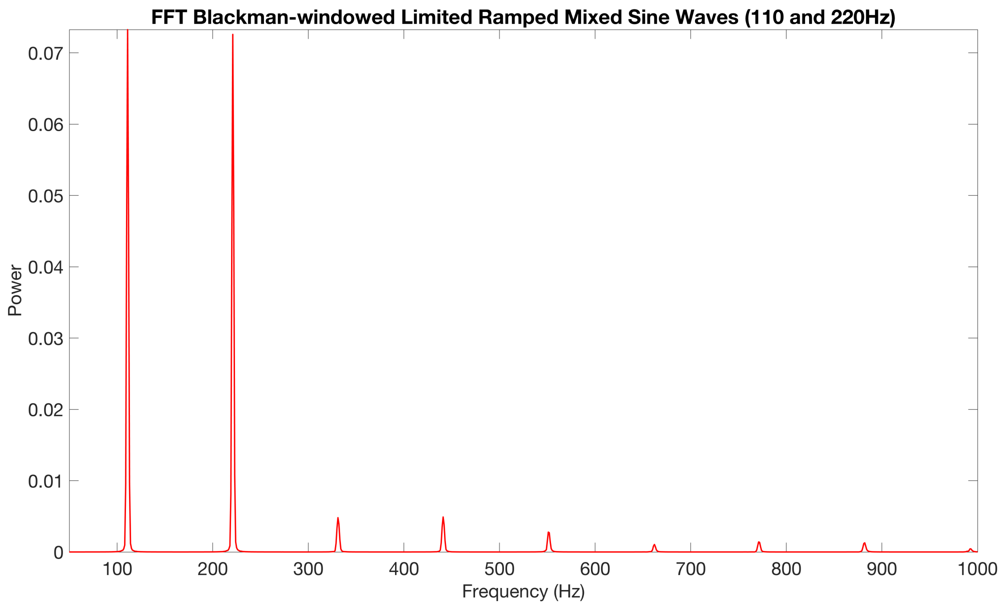


Figure 8.13: Test 2A - FFT of XMOD HRL signals demonstrating an alternative threshold (light DRC).

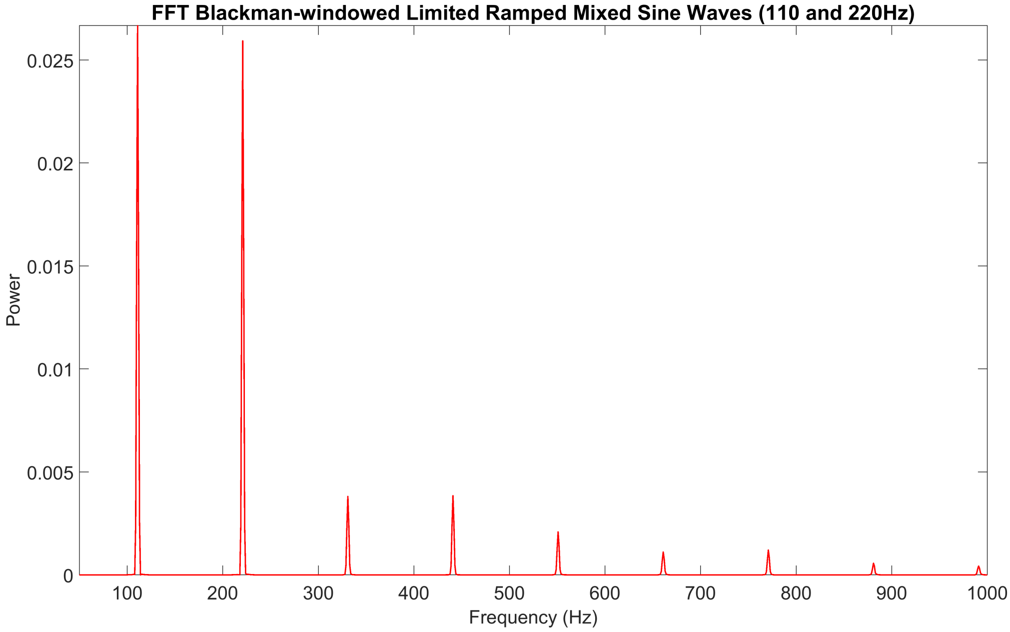


Figure 8.14: Test 2A - (default threshold) FFT of the XMOD HRL signals post DRC application (HL DRC).

Applying two different DRC magnitudes (light, and heavy) to HUR sine waves (Test 2B) (Fig. 8.15 and 8.16 respectively) again demonstrates the progressive nature of increased DRC. Harmonic distortion components are present, like test 2A (i.e., at 330 Hz, 440 Hz, 550 Hz, 660 Hz, 770 Hz, and so on) as well as the addition of IMD products (i.e., sum and difference components discussed in section 3.4). It is important to observe in all of the XMOD FFT's that the 196 Hz and 220 Hz signals are more attenuated than the 110 Hz signal.

Previous investigations by this researcher have highlighted an anomaly not evident in the preceding FFTs in which low-frequency content modulating under DRC is subject to higher attenuation than other simultaneous signals (Campbell, Toulson, and Paterson, 2010; Ward, and Campbell, 2010; Campbell, 2012; Campbell, Paterson, and Toulson, 2014; Campbell, 2014). Figure 8.17 shows the lowest frequency modulated under DRC, 55 Hz (A0), is more attenuated than the HRL 110 Hz signal. However, the effect is diminishing for HUR signals as observed in Fig. 8.18, 73 Hz (D1).

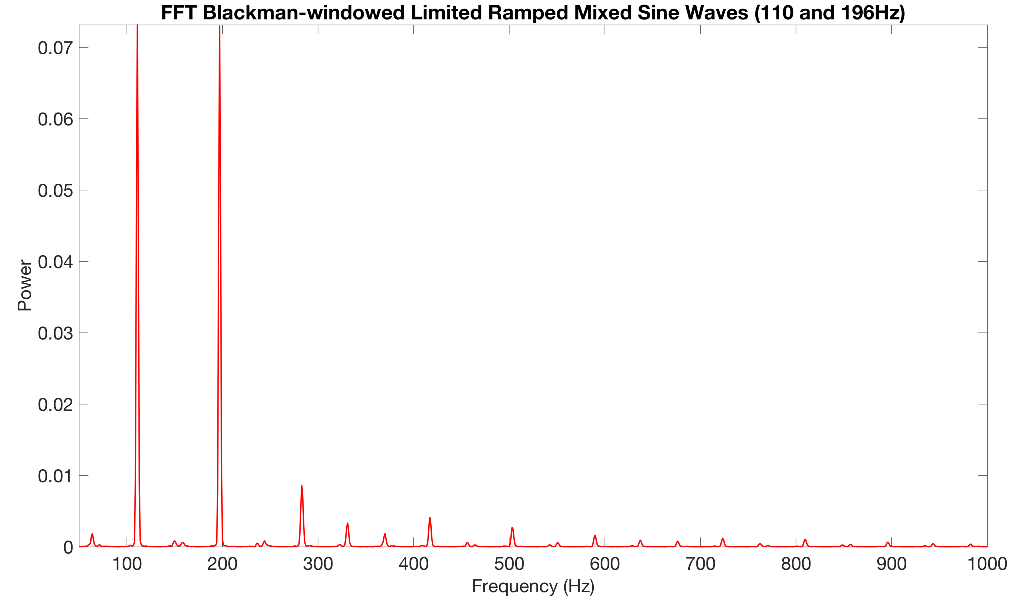


Figure 8.15: Test 2B - FFT of XMOD HUR signals demonstrating an alternative threshold (light DRC).

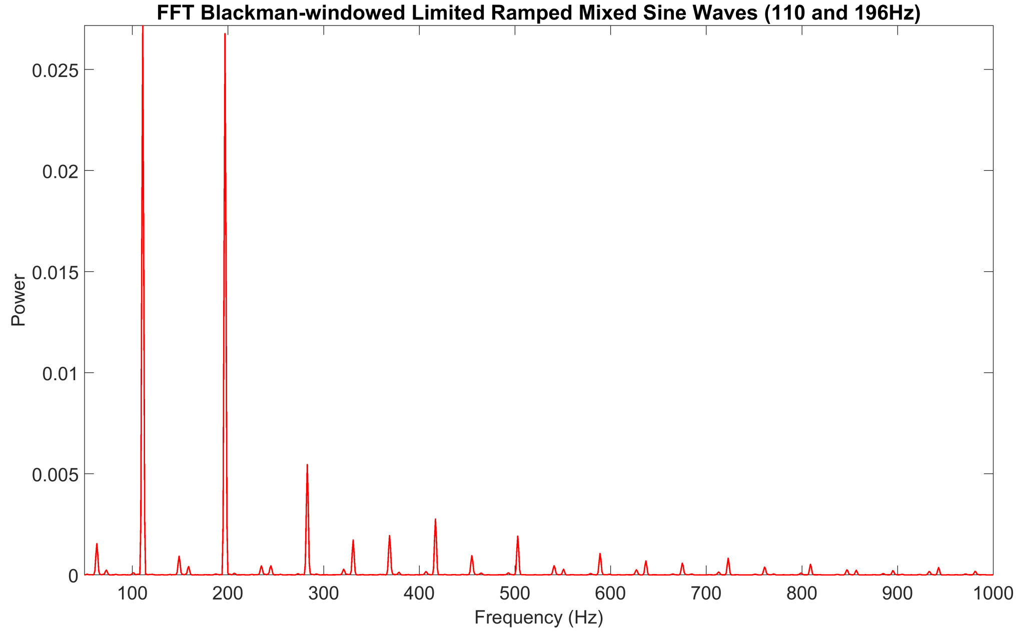


Figure 8.16: Test 2B - the default threshold (HL), FFT of the XMOD HUR signals post DRC application.

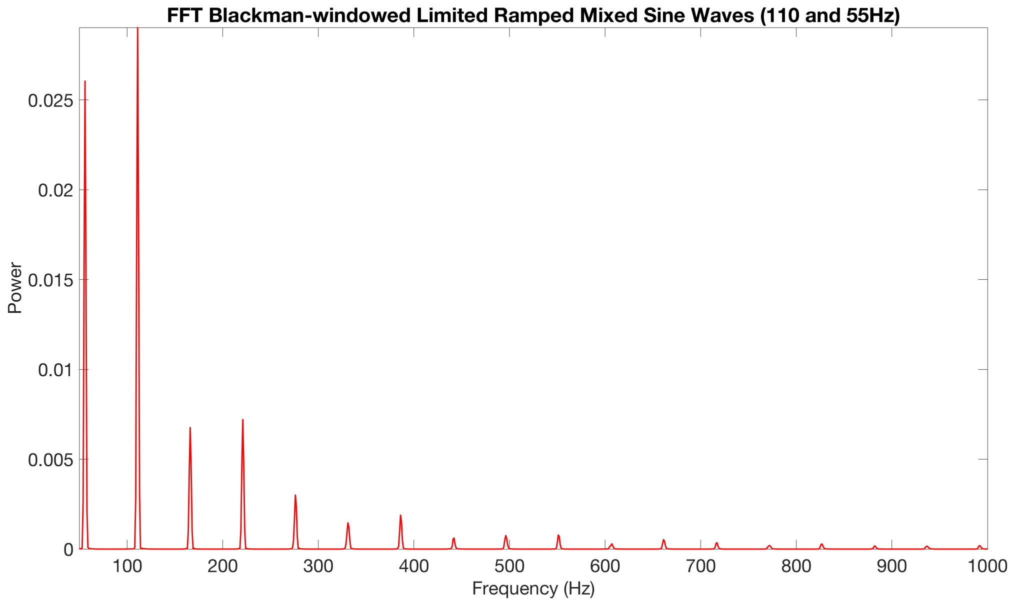


Figure 8.17: Test 2B - FFT of XMOD HRL signals demonstrating an alternative low frequency (55 Hz).

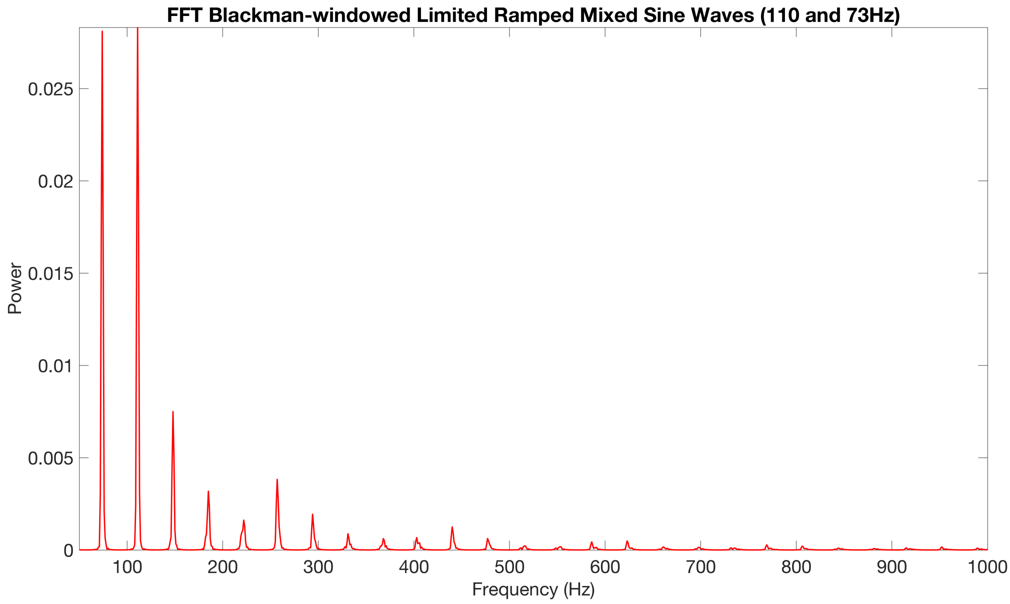


Figure 8.18: Test 2B - FFT of XMOD HUR signals demonstrating an alternative low frequency (73 Hz).

Visual analysis of the plots for the XMOD NC (Fig. 8.19) signals show relatively similar results to those of Test 1. The loudness and RMS amplitude for the HUR signals are a steadier state than the HRL signal.

The plots for the XMOD HL (Fig. 8.20) signals also show relatively similar results of DRC shaping the waveforms, increasing the loudness, and RMS amplitude for both HRL and HUR. The transform plots are similar to the HL plots for the single sine wave, fluctuating in sympathy with the signal, with the most deflection at the beginning of the ramp and the least at the very end. The RMS amplitude for both HRL and HUR is fast to rise and maintain similar deflections.

The LFC transform of the HUR signal fluctuates much more than the HRL signal, indicating more distortion.

The ABS diff. between the mean amplitude of the NC stimulus and the HRL stimulus is 4.64 dB for and 24.26 dB for the HUR stimulus. It is not clear why the mean of the HRL test signal is slightly higher, and much lower for HUR, compared to NC; perhaps the greater modulation between the HUR signals compared to the HRL signals means more power is lost to the distortion components, as shown in Fig 8.16.

The MAE of both XMOD stimuli (HRL and HUR) (0.15) indicates that applying DRC to a cross modulating sine waves does not alter the amplitude of either signal notably and that the performance of DRC applied to HRL and HUR signals is similar.

XMOD	**HRL Stimuli**		**HUR Stimuli**	
	NC	***HL***	***NC***	***HL***
n (samples)	44101.00	44101.00	44101.00	44101.00
Mean (dB)	-59.29	-54.65	-58.94	-83.20
ABS diff. (dB)	4.64		24.26	
MAE	0.15		0.15	

Table 8.3: Statistical comparison of XMOD HRL and HUR signals.

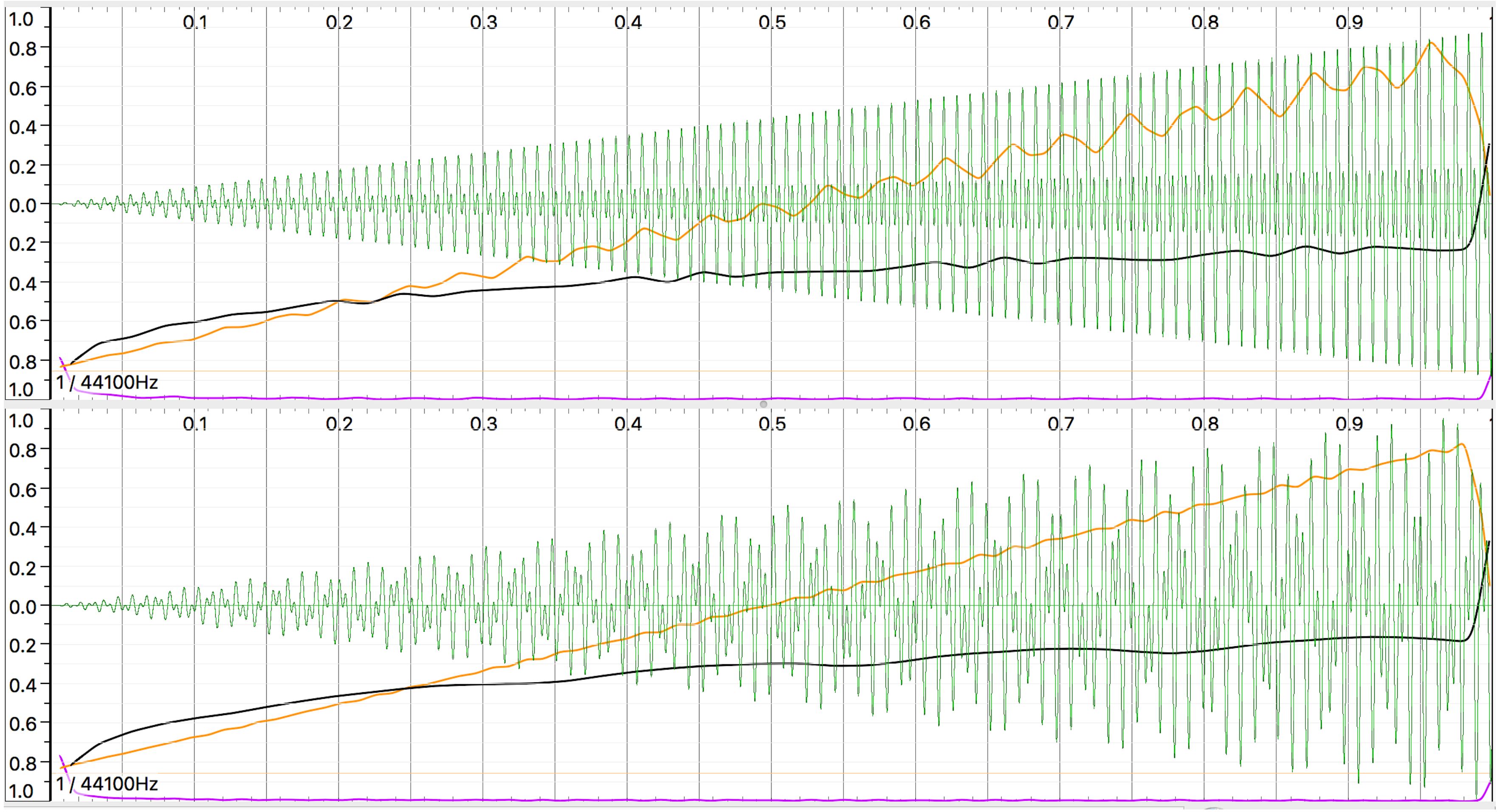


Figure 8.19: Amplitude vs. time (ms) plot showing XMOD signals; HRL NC in the top pane and HUR NC in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

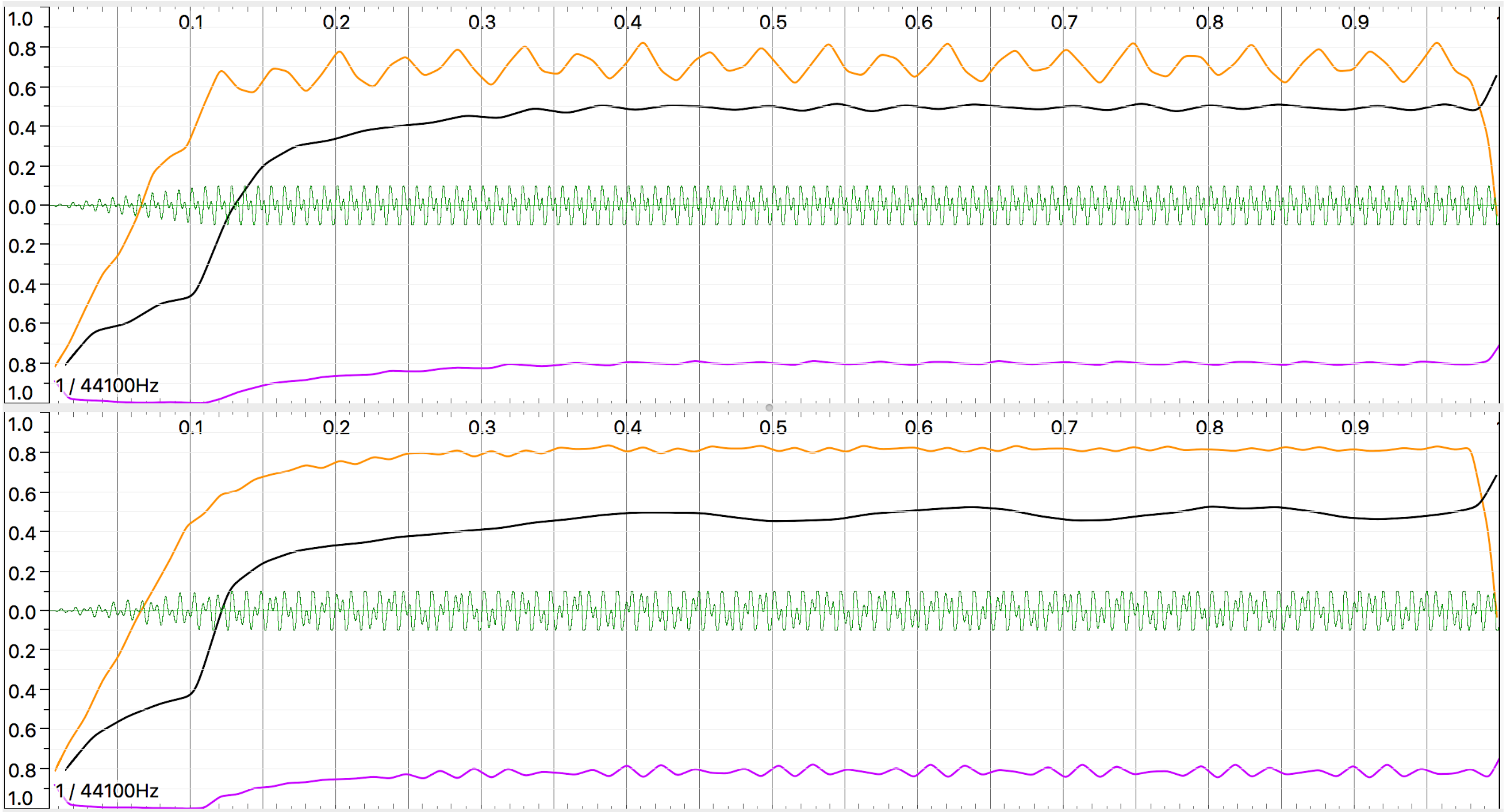


Figure 8.20: Amplitude vs. time (ms) plot showing XMOD signals; signals HRL HL in the top pane and HUR HL in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

Table 8.4 shows the statistical comparison of BSLoud, RMS amplitude, and LFC for the XMOD stimuli. The increase in psychoacoustically-informed loudness for each XMOD comparison is of 2.22 dB for the HRL HL stimulus and 1.92 dB for the HUR HL stimulus. This is also reflected by the greater MAE of 0.54 units for the HRL stimulus and 0.48 units for the HUR stimulus.

RMS has an absolute difference of 12 dB for both configurations. This is reflected in a MAE of 0.19 bels for both tests.

The LFC mean frequency of HRL has an upward shift of about 39 Hz and the HUR has an upward frequency shift of about 54 Hz. This is reinforced with an MAE of 40 Hz for HRL and 54 for HUR stimuli compared to NC.

XMOD	**BSLoud**			
	HRL Stimuli		**HUR Stimuli**	
	NC	***HL***	***NC***	***HL***
n (samples)	44.00	44.00	44.00	44.00
Mean (dB)	10.59	12.80	10.87	12.78
ABS Diff. (dB)	2.22		1.92	
MAE	0.54		0.48	
	RMS Amplitude			
n (samples)	87.00	87.00	87.00	87.00
Mean (dB)	-12.05	-24.58	-12.04	-24.29
ABS Diff. (dB)	12.53		12.24	
MAE	0.19		0.19	
	LFC			
n (samples)	88.00	88.00	88.00	88.00
Mean (Hz)	154.48	193.61	148.06	202.50
ABS Diff. (Hz)	39.13		54.44	
MAE	40.56		54.44	

Table 8.4: Quantification of XMOD signals, configured as both HRL and HUR. ABS diff. is the absolute differential between means of the control and test stimuli.

8.3.3 Test 3: DRC (INDEP)

Test 3 has two parts: 3A applies DRC to two INDEP signals constructed from HRL frequencies (110 Hz and 220 Hz); 3B applies DRC to two INDEP before mixing constructed from HUR frequencies (110 Hz and 196 Hz).

Visual analysis of the HL FFTs for INDEP HRL (Fig. 8.21) and HUR (Fig. 8.22) stimuli suggests that mixing the signals after applying DRC may reduce distortion artefacts considerably, regardless of the signals frequencies being HRL or HUR. By mixing the signals post DRC, each signal modulates independently, and therefore only odd integer harmonic components related to the fundamental frequencies are produced (i.e., at 330 Hz, 550 Hz, 770 Hz, and so on; plus 660 Hz, 1100 Hz, 1540 Hz, and so on). There seems to be a minute amplitude reduction of the 220 Hz (HRL) and 196 Hz (HUR) signals. At this point, there is no explanation for this.

Visual analysis of the plots for the INDEP NC (Fig. 8.23) signals show relatively similar results to those of Test 1 and Test 2. Figure 8.24 exhibits lower levels for loudness, RMS and LFC than Test 2 (Fig. 8.18), much more comparable to the plots produced for Test 1 (Fig. 8.11).

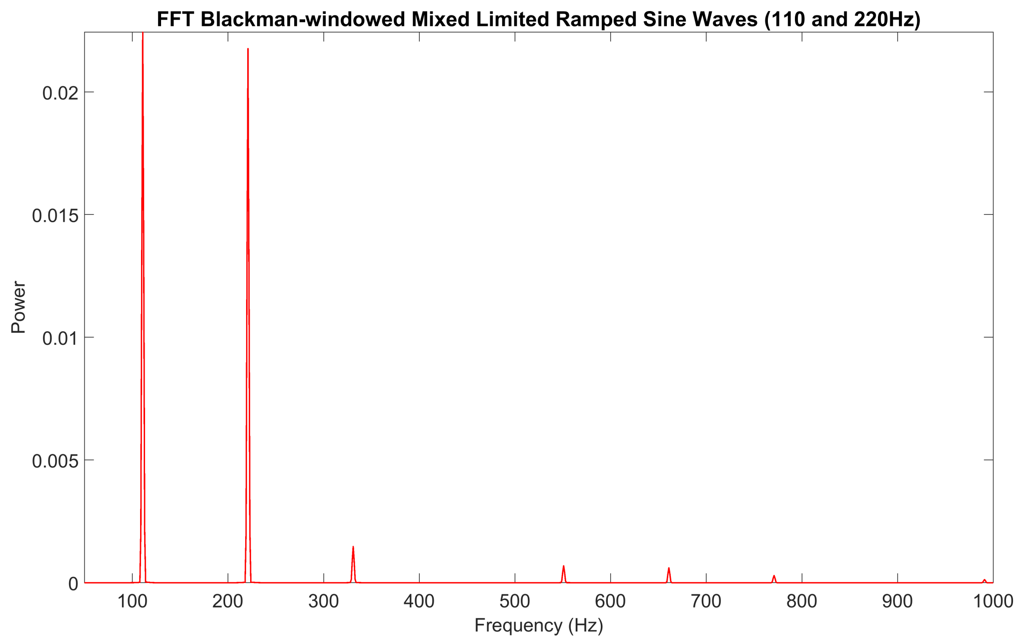


Figure 8.21: Test 3A - FFT of the IDEP HRL signals post DRC application.

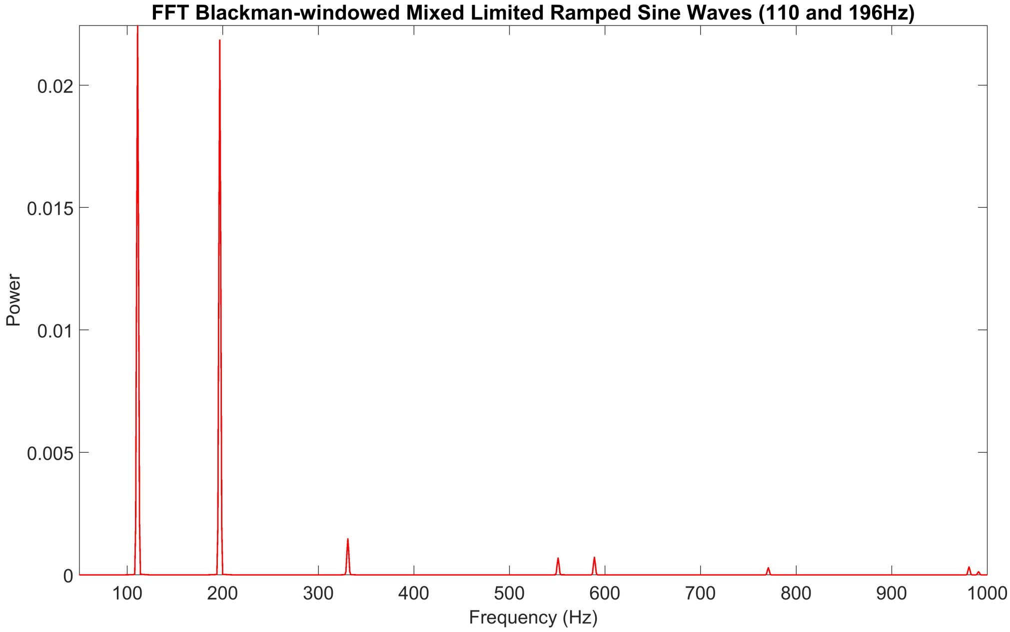


Figure 8.22: Test 3B - FFT of the IDEP HUR signals post DRC application.

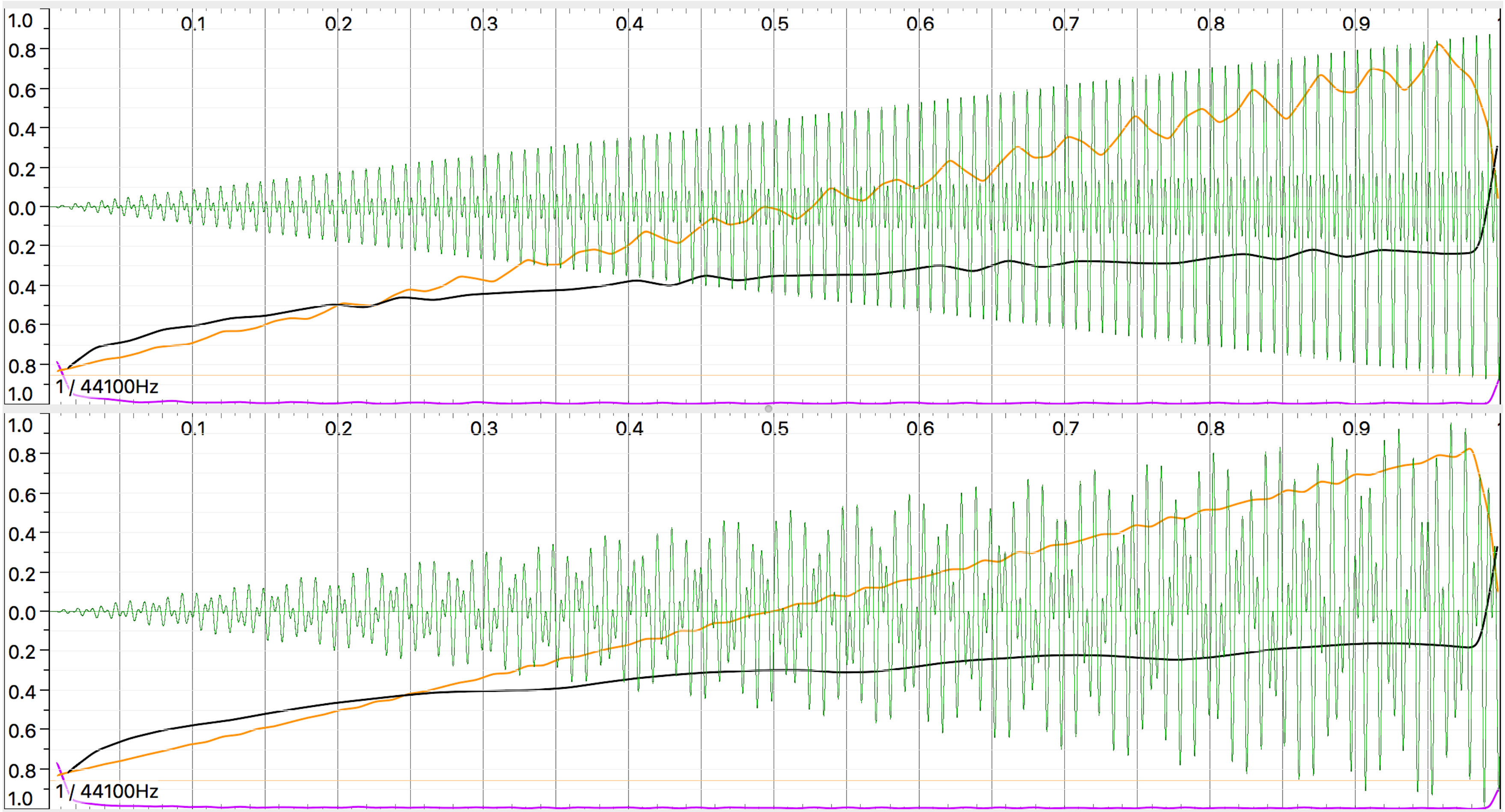


Figure 8.23: Amplitude vs. time (ms) plot showing INDEP signals; HRL NC in the top pane and HUR NC in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

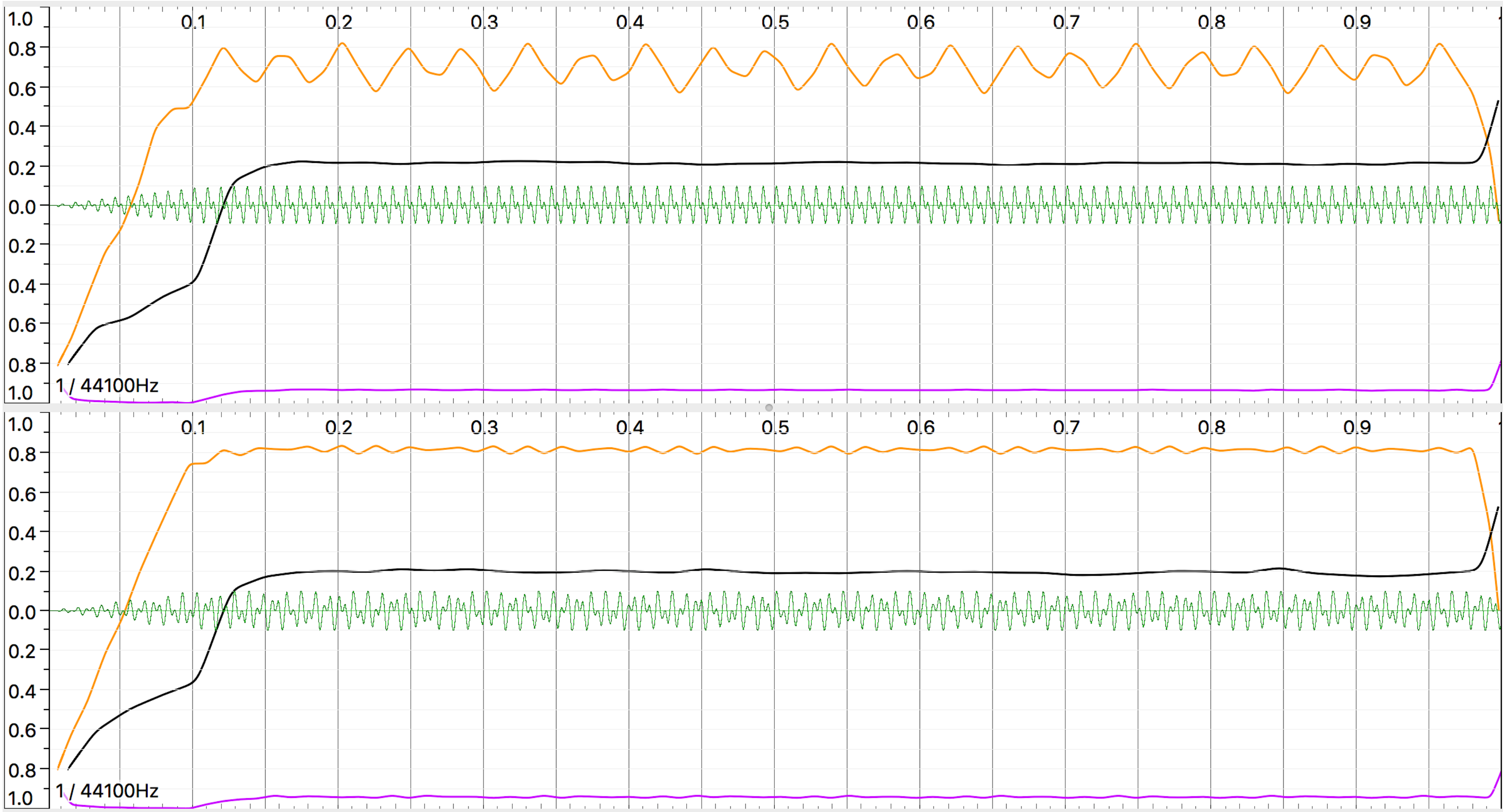


Figure 8.24: Amplitude vs. time (ms) plot showing INDEP signals; HRL HL in the top pane and HUR HL in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

The INDEP ABS diff. between the mean amplitude of the NC stimulus and test signals are 19.5 dB for both the HRL and HUR comparisons, much more consistent than for the XMOD stimuli and in line with the 19.4 dB mean amplitude of the single sine.

The MAE comparison of NC and HL of both XMOD stimuli (HRL and HUR) (0.16) indicates that applying DRC to independently modulating sine waves does not introduce significant error to either signal and that the performance of DRC applied to HRL and HUR signals is similar.

INDEP	**HRL Stimuli**		**HUR Stimuli**	
	NC	***HL***	***NC***	***HL***
n (samples)	44101	44101	44101	44101
Mean (dB)	-59.29	-78.79	-58.94	-78.43
ABS Diff. (dB)	19.50		19.49	
MAE	0.16		0.16	

Table 8.5: Statistical comparison of the INDEP signals with and without DRC. ABS diff. is the absolute differential between means of the control and test stimuli.

Table 8.6 shows the statistical comparison of BSLoud, RMS amplitude, and LFC for the INDEP stimuli. The BSLoud increase in psychoacoustically-informed loudness for each INDEP comparisons is 0.74 dB for the HRL HL stimulus and 0.32 dB for the HUR HL stimulus. This is also reflected by the greater MAE of 0.22 units for the HRL stimulus and 0.17 units for the HUR stimulus.

RMS HL absolute difference from NC of about 14 dB; 2 dB more than the XMOD test stimuli. This is reflected in a MAE of 0.20 Bels for both tests.

The LFC mean frequency of HRL has an upward shift of about 19.74 Hz (about 19 Hz less than XMOD) and the HUR has an upward frequency shift of about 17.48 (about 37 Hz less than XMOD). This is reinforced with a MAE of 19.74 Hz for HRL and 17.49 for HUR stimuli compared to NC.

INDEP	BSLoud			
	HRL Signals		HUR Signals	
	NC	***HL***	***NC***	***HL***
n (samples)	44.00	44.00	44.00	44.00
Mean (dB)	10.59	11.33	10.87	11.19
ABS Diff. (dB)	0.74		0.32	
MAE	0.22		0.17	
	RMS amplitude			
n (samples)	87.00	87.00	87.00	87.00
Mean (dB)	-12.05	-26.08	-12.04	-26.05
ABS Diff. (dB)	14.03		14.01	
MAE	0.20		0.20	
	LFC			
n (samples)	88.00	88.00	88.00	88.00
Mean (Hz)	154.48	174.21	148.06	165.54
ABS Diff. (Hz)	19.74		17.48	
MAE	19.74		17.49	

Table 8.6: Quantification of INDEP signals, configured as both HRL (Figure 3.3.4) and HUR). The red indicates stimuli that significantly differ from their corresponding control stimulus. ABS diff. is the absolute differential between means of the control and test stimuli.

8.3.4 Final Statistical Comparison of Sine Test Results

Final analysis of the overall sine wave findings shows that there is a pattern to the findings for the application of DRC to sine waves, mixed or not. Table 8.7 ranks the experimental configurations according to their MAE compared to their corresponding NC stimulus. The table shows that signals combined as INDEP have the least error when compared with NC, followed by XMOD and then the single HL sine wave. These findings also show lower MAE for HUR stimuli compared to HRL stimuli.

Rank	DRC Configuration	ABS Difference	MAE
Waveform Similarity (ranked low to high MAE)			
1	INDEP HUR	19.49	0.160
2	INDEP HRL	19.50	0.163
3	XMOD HUR	24.26	0.151
4	XMOD HRL	4.64	0.152
5	Sine HL	19.37	0.254

Table 8.7: Ranked DRC configurations according to their MAE comparison to NC.

Table 8.8 shows the three measures (loudness, RMS and LFC) ranked again from lowest MAE to highest as compared to their corresponding NC stimulus. These combined results indicate that INDEP has the lowest MAE in every test followed by XMOD and Sine HL in varied ordering.

Rank	DRC Configuration	ABS Difference	MAE
Loudness (ranked low to high MAE)			
1	INDEP HUR	0.32	0.172
2	INDEP HRL	0.74	0.213
3	XMOD HUR	1.92	0.479
4	Sine HL	0.95	0.482
5	XMOD HRL	2.22	0.541
RMS Amplitude (ranked low to high MAE)			
1	INDEP HRL	14.03	0.198
2	INDEP HUR	14.01	0.198
3	XMOD HUR	12.24	0.187
4	XMOD HRL	12.53	0.189
5	Sine HL	13.99	0.317
Log Frequency Centroid (ranked low to high MAE)			
1	INDEP HUR	17.48	17.486
2	INDEP HRL	19.74	19.741
3	Sine HL	21.09	21.092
4	XMOD HRL	39.13	40.557
5	XMOD HUR	54.44	54.441

Table 8.8 Ranked DRC configurations according to their MAE comparison to NC.

8.4 Method (Analysis using Pro Tools)

The second part of this chapter is an experiment designed to test the effects of DRC on musical signals. These experiments are intended to build on what we learned about the DRC from the previous sine wave experiments to test the potential similarities, and use that knowledge to increase our understanding about DRC applied to musical signals.

8.4.1 Apparatus

Stimuli were recorded using various microphones with an appropriate gain structure. The microphones fed the inputs of a Yamaha DM1000 digital mixing desk. Signal output from the mixing desk was via the FireWire 800 interface and lead connected to an Apple MacBook Pro. Audio recording was via Apple Logic Pro software, exported (not bounced) prior to summation to Avid Pro Tools software for mixing. The levels of each track were locked to their resultant amplitude by bouncing each track out separately, and imported into another Pro Tools project with each track set to unity gain. After, experimental application of DRC used the peak-sensing Pro Tools native Dynamics III Compressor/Limiter. Analysis of the resultant musical stimuli used Sonic Visualiser (Cannam, Landone, and Sandler, 2010) using Libxtract: Loudness; RMS Amplitude (Bullock, 2007); Spectral Centroid (Cannam, Landone, and Sandler, 2010) and Microsoft Excel.

8.4.2 Stimuli

Overdub style recording (as opposed to recording instruments played simultaneously) of a musical ensemble produced stimuli subsequently used for DRC experimentation. The overdubbing method requires recording of instruments at separate times, building up layers of performances to form a complete musical piece. This approach avoids issues associated with spillage between instruments, therefore aiding the mixing process (Borwick, 1996) and simplifies analysis of the stimuli produced.

The five-piece pop ensemble Bijoumiyo from Cambridge (UK) performed their piece, ‘Uncover My Eyes’ for use in the production of this thesis. Tracking was over two days in

Anglia Ruskin University's Helmore Recital Hall. The author of this thesis produced and engineered the song with assistance from undergraduate students. Selection and placement of microphones were with great care, ensuring zero usage of additional signal processing during tracking or mixing was required. The monophonic recording and mix utilised (3:34-minutes long) sampling rates were at 44.1 kHz, 16-bit sampling. Resultant stimuli comprise thirteen tracks (bass drum, drum overhead, snare drum, tom-tom 1, tom-tom 2, congas, bass guitar, acoustic guitar, trumpet, grand piano, Vox 1, Vox 2, Vox 3) without DRC, equalisation, or any other post-production effects, mitigating against the possibility of unwanted artefacts.

Every DRC configuration was normalised to -15 dB RMS prior to DRC application to ensure constancy in the application of DRC. The DRC used for this part of the experiment is the Pro Tools native Dyn3. The Dyn3 compressor/limiter utilises look-ahead functionality (Avid, 2013), though the manufacturer does not specify the signal delay time. The parameters of the attack and release times employed the fastest possible setting for the device. The ratio is the highest available, and the threshold set to the approximately 3 dB below RMS (cf. Stone et al., 2009; Katz, 2007). Aggressive settings were used intentionally to attempt to achieve a full-sum outcome similar to CD version of the Metallica track discussed in section 2.3.

Next, Pro Tools was configured for applying DRC at one of three points (Fig. 8.25) in the signal path: 1) 'Track', having DRC on each track only; 2) 'Subgroup', having DRC on each of three subgroups (consisting of a drum subgroup, other (instrument) subgroup, and vocal subgroup, only; 3) 'Full-sum', having DRC on the sum of the three subgroups only.

The configuration of each track, subgroup, and full-sum incorporated a Pro Tools Dynamics III Compressor/Limiter, simultaneously, in bypass mode as required. This configuration ensured that the signal path remained unaltered for each stimulus generation. Activation of the appropriate DRC component produced each of the three DRC configurations described above and deactivated otherwise. The makeup gain was not applied, nor were the signals normalised post-DRC. The reason for this choice was so that signal analysis would be as output. In ‘real-world’ music production, the reduction in signal RMS via DRC allows makeup gain to raise the amplitude of the signal with reduced dynamics, which in turn increases the perceived loudness compared to the more dynamic signal.

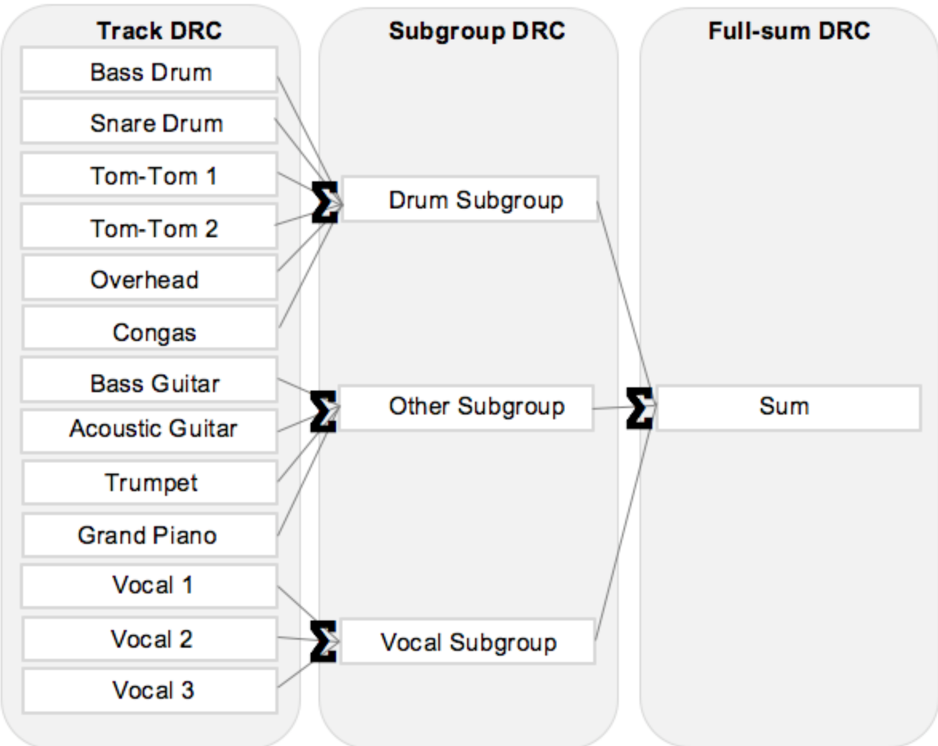


Figure 8.25: Signal path for grouped stimulus production.

8.4.3 Procedure

This section employs two tests to quantify the effect of DRC on musical signals, each described below.

Test 1 is a statistical comparison of musical stimuli configured as NC and HL signals to determine if an in-depth examination of DRC musical stimuli is warranted. The waveforms are not compared directly due to the large amount of data, and therefore comparisons are made with using the three transforms utilised in the sine experiment (i.e., BSLoud; RMS amplitude; and LFC).

Some points of interest from the sine wave tests are loudness modulations relative to the amplitude of the signals; alterations of harmonic content, i.e. the reduction associated with the fundamental frequency of sine waves under DRC; DC offset; large absolute differentials between signal means (ABS Diff.); mean absolute error (MAE).

Data gathered from the three transforms is exported as text documents and imported into an Excel spreadsheet for statistical analysis using mean, ABS Diff., and MAE. Other Sonic Visualiser (SV) analysis includes variations in the spectrograms of the stimuli.

Test 2 is a comparison of the signals summed post-DRC (XMOD) and pre-DRC (INDEP). Stimuli included NC, and DRC applied on discrete tracks, subgroup, and full-sum. Analysis of the DRC configurations utilised SV for further insight into the effects of DRC. Again, analysis of Bark scale loudness (BSLoud), RMS amplitude, and log frequency centroid (LFC) was made using SV transforms.

Data gathered from the three transforms was exported as text documents to Excel spreadsheets for statistical analysis using mean, ABS Diff. between means, and MAE.

8.5 Pro Tools Results and Discussion

8.5.1 Test 1: Comparison of NC to DRC musical signals

Figure 8.26 shows two waveforms of the test track; Bijoumiyo's 'Uncover My Eyes'. The top waveform has no-compression (NC) and the bottom has heavy limiting DRC (HL). The HL track has similar DRC ('cut grass' appearance) to the Metallica track from section 2.3. The plot consists of a seven-second excerpt of the stimuli, and plots of BSLoud, RMS amplitude, and LFC.

Interestingly, the loudness transform generally follows the envelope of NC. Roughly the same loudness envelope and signal level are evident for HL, though altered in places. The RMS transform has a much higher average for HL than for NC, with more significant variation in places.

The statistical analysis (Table 8.9) shows relatively similar results comparing the mean and ABS Diff. for HL RMS, BSLoud and LFC to NC. However, MAE for loudness does reflect the presence of distortions realised in the sine wave tests. What this tells us is that the measurement envelopes are likely almost identical but significantly removed from each other in some way, as verified in figure 8.27. This is most reflected by the exaggeration of the RMS envelope of HL as compared with NC. LFC appears very correlated for HL and NC, with only subtle differences evident, which would follow the observations of the sine waves.

Further analysis of the RMS and BSLoud envelopes in Fig. 8.26 and 8.27 would indicate the apparent reduction of the transient portions of the signals with an increased perceptible loudness of the portions which follow the transient which would include instrument tonality, direct reflections and reverberant information, features not present in sine waves.

.	**RMS Amplitude**		**BSLoud**		**LFC**	
	UME NC	***UME HL***	***UME NC***	***UME HL***	***UME NC***	***UME HL***
n (samples)	18605	18605	9303	9303	18606	18606
Mean	-26.25 dB	-26.67 dB	16.02 dB	16.21 dB	63.55 Hz	63.66 Hz
ABS Diff.	0.42 dB		0.19 dB		0.11 Hz	
MAE	0.01		0.41		52.91	

Table. 8.9 Quantification the piece of music with and without DRC.

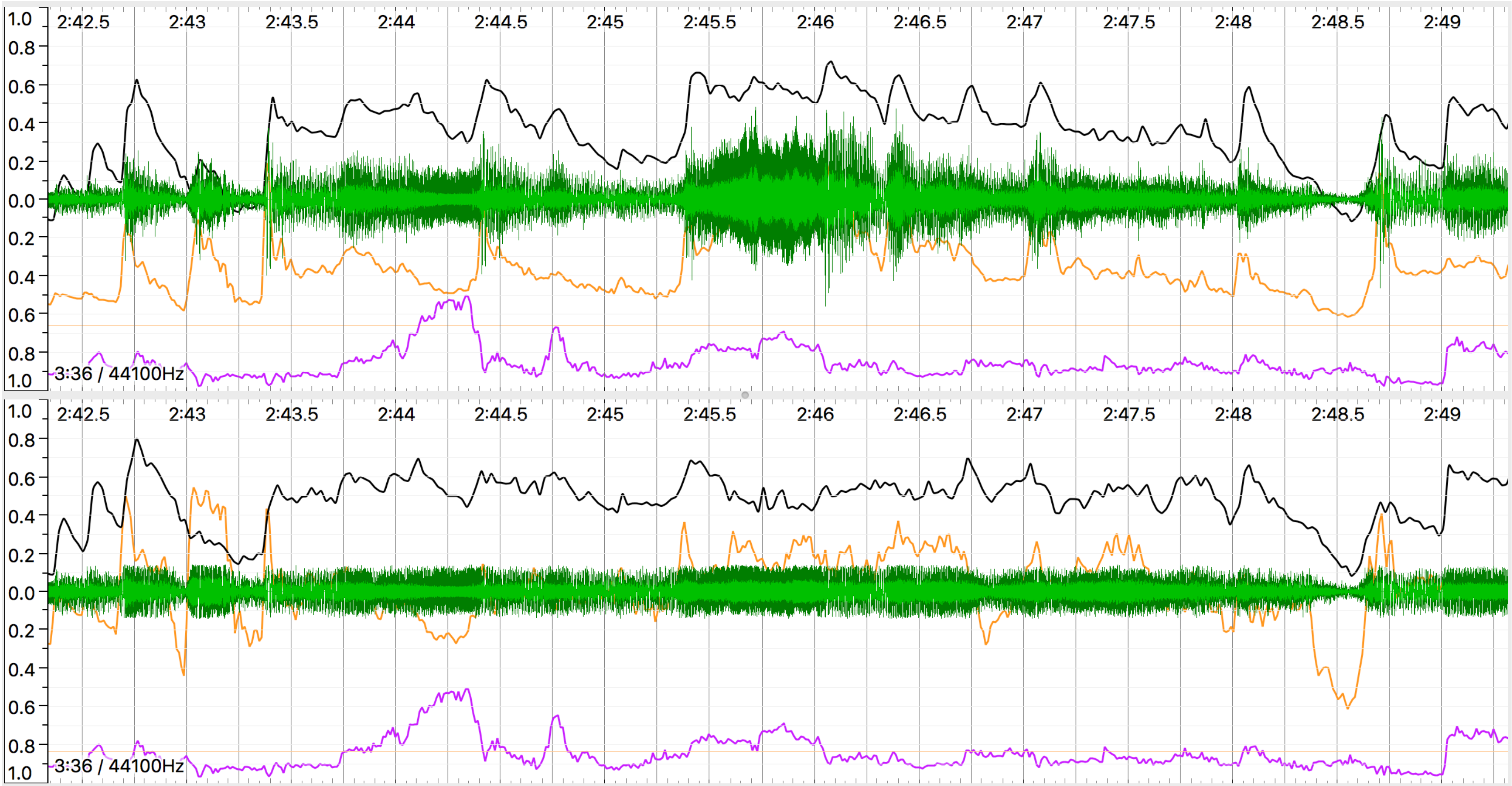


Figure 8.26: Amplitude vs. time (min:sec) plot showing music signals; UME NC in the top pane and UME HL in the bottom pane. Both signal waveforms are shown in green, loudness transforms in black, RMS amplitude transforms in orange, and the LFC transforms in magenta.

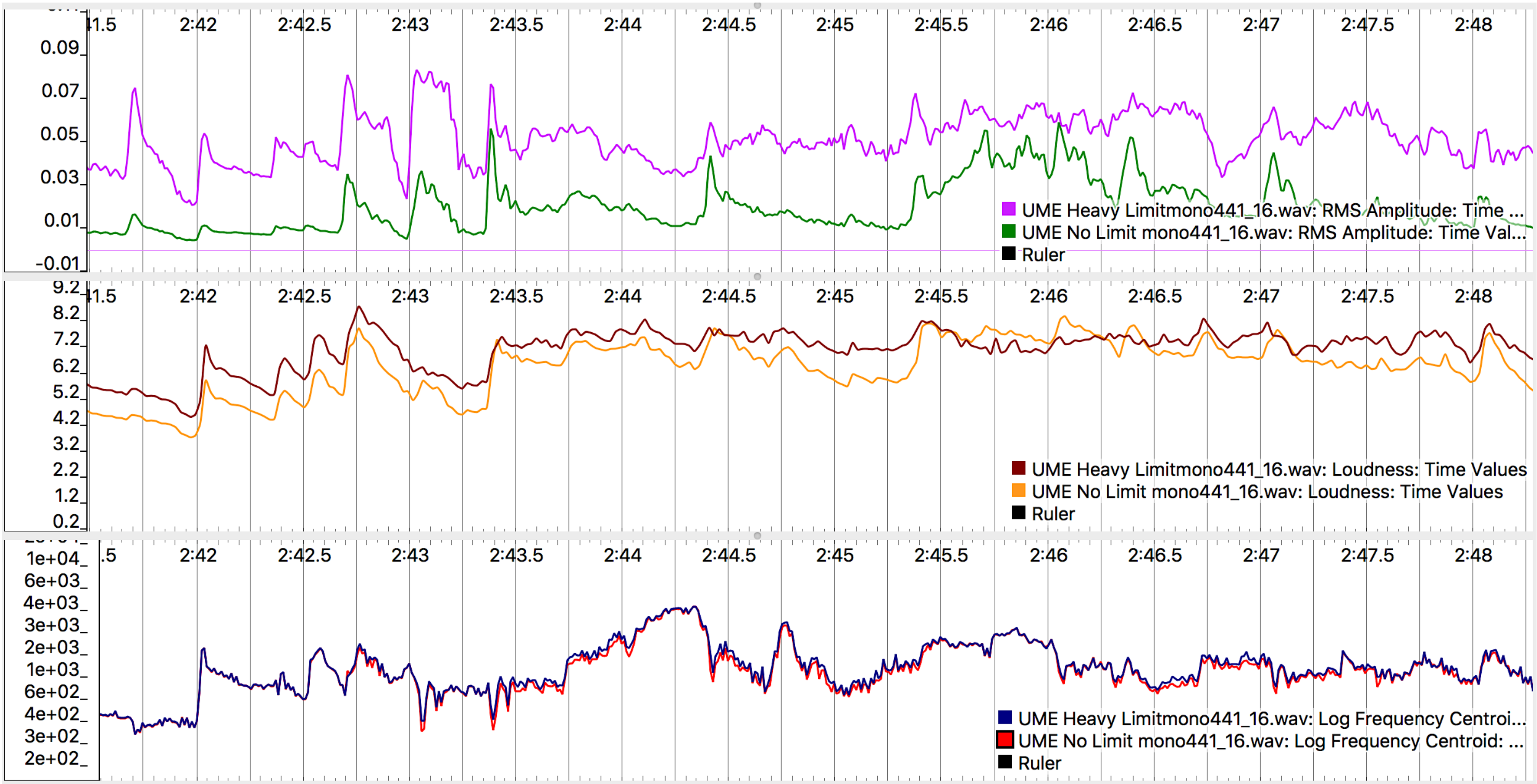


Figure 8.27: A like for like comparison of UME NC and HL; the top pane shows RMS amplitude (dB) vs. time (min:sec); the middle pane is BSLoud amplitude (dB) vs. time (min:sec); the bottom pane is LFC frequency (Hz) vs. time (min:sec).

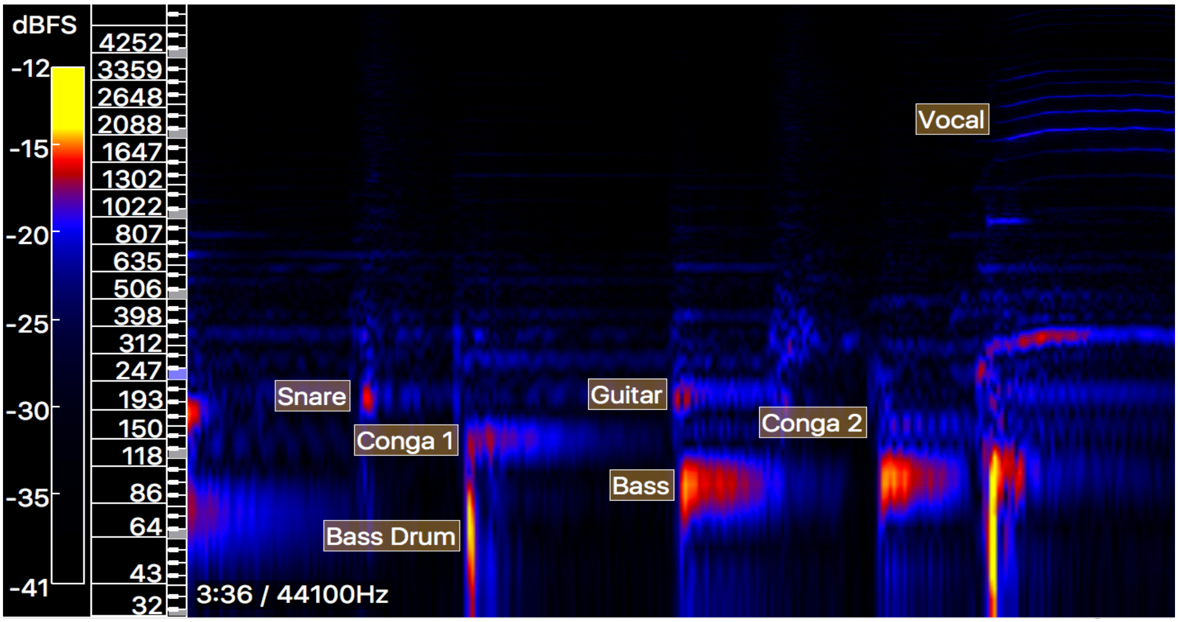


Figure 8.28: The spectrogram indicating magnitude (dBFS) and frequency (Hz) vs. time (ms) indicating typical instrument envelopes and amplitudes. The colours represent amplitude shown in the meter on the left, next to the frequency scale.

Figure 8.28 illustrates how to read the spectrogram, showing the typical instrument envelopes, positions on within the frequency spectrum and typical amplitudes encountered in this section. Note that the scale of the spectrogram may change, depending on the desired focus.

The spectral analysis of the passage illustrated in Fig. 8.28 reveals some points of interest for further investigation, depicted as zones 1-3 which will be analysed further respectfully following the brief introduction to significance the each zone. For instance, zone 1 (Fig. 8.28) compares amplitude variations of single instruments (bass drum and bass guitar) that appear rearranged dynamically (comparing the top NC pane with the lower HL pane). As each of the bass drum notes coincide with the bass guitar notes, alterations introduced by DRC are evident which may or may not be dependent interaction between each instrument under DRC. This rearrangement of the amplitude dynamics can have the effect of altering the rhythmic structure, indicative of the ‘feel’ of the piece (Campbell, 2012; Toulson and Campbell 2009; Campbell, Paterson, and Toulson, 2010).

Zone 2 (Fig. 8.28) potentially shows rearrangement of amplitude within the harmonic structure and thus the tonality of individual signals (Ward and Campbell, 2010; Toulson and Campbell 2009). This is similar to the effect briefly addressed in zone 1, but in this case DRC may be affecting a single instrument. The harmonic alterations here may be analogous with the observations of the HRL sine wave tests in which the amplitude balance between the individual sine components was altered depending on frequency content under DRC.

Zone 3 (Fig. 8.28) shows one instrument affecting the output of another while under DRC. Here, the bass guitar appears to reduce the amplitude of the bass drum. Although potentially related to the zone 1 observation, in music production this same type of effect is an intentional or unintentional consequence of DRC, earning particular attention, described it as ‘pumping’ and ‘breathing’ (Case, 2007; Izhaki, 2010).

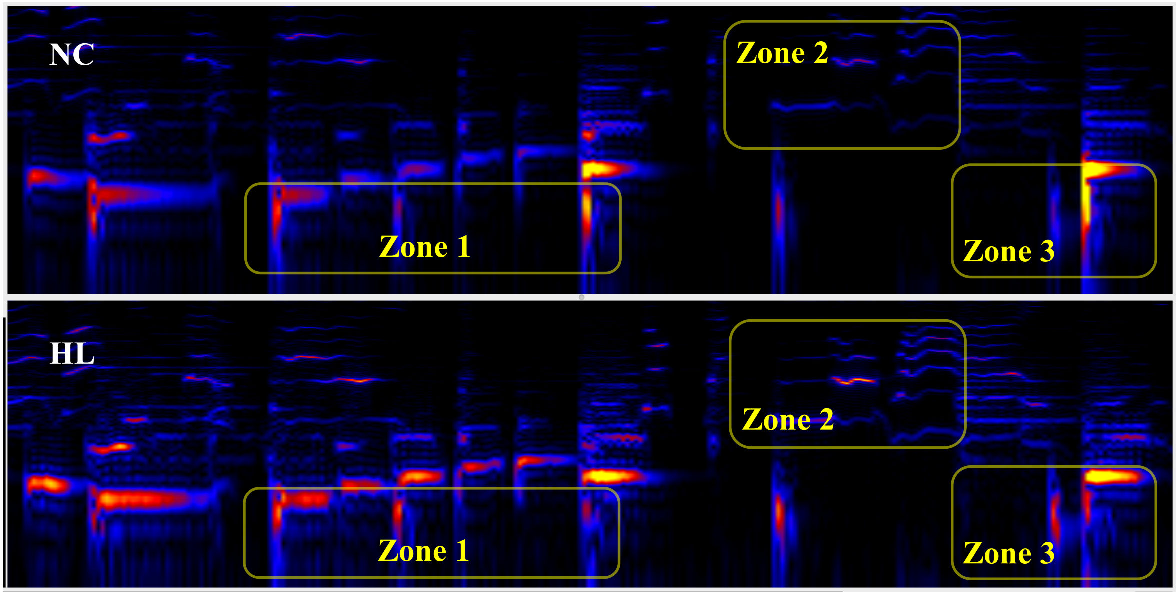


Figure 8.29 The spectrogram (amplitude and frequency vs. time) shows potential problem areas within a short window of the UME test audio. The top pane shows UME NC and the bottom pane UME HL. Zone 1: dynamical changes; zone 2: harmonic changes; zone 3: instrument interaction changes.

A) Zone 1: Investigation of Dynamical Changes

The first point of interest is the drum dynamics and transient response under DRC. Figure 8.30 spectrally 'maps out' the dynamics of a seven-second segment of bass drum notes, annotated with their corresponding amplitudes. The figure also measures the bass guitar notes to assess the interaction between two instruments when under DRC. Also included is a plot of the RMS amplitude envelope for both configurations.

Note that the bass drum notes have a recognisable dynamic pattern, produced by alternating the impact velocity: (illustrated using font alterations such as **BD** for hard hit and *BD* for a softer hit etc.) *BD1*, **BD2**, *BD3*, **BD4**, *BD5*, *bd6*, **BD7**. Normally these dynamics correspond to the timing and feel of the piece of music, and often from interactions with other instruments in the performance. The best players use dynamic variations for instrumental expression in any creative music.

The amplitude variations of the signal in the HL window illustrate alteration of the dynamic pattern: *BD1*, **BD2**, **BD3**, *BD4*, **BD5**, **BD6**, bd7. A good illustration of this is the inversion of BD6 and BD7 (the first note is a ghost note in NC) ending the segment. The RMS signal envelope seems to demonstrate that a bass drum signal contains most energy in the transient, indicated from the sharp spikes in RMS coinciding at the bass drum transients. When the bass guitar signal and the bass drum transients correspond under DRC, the bass drum is attenuated. When not interacting with the bass notes the bass drum appears to gain amplitude, as would be expected behaviour of the DRC.

Figure 8.31 illustrates BSLoud, RMS amplitude and LFC; each compared statistically in Table 8.10. The mean and ABS diff. for zone 1 RMS amplitude and BSLoud are again pretty similar to NC, whereas HL LFC is about 10 Hz above NC. It is interesting that the ABS Diff. for LFC is greater for this section than the test 1 section, yet the LFC MAE comparison of UCM NC and HL is lower here; perhaps this disparity demonstrates the alterations to the signal structure this part of the test is analysing. Close inspection of the excerpt in Fig. 8.30 reveals that the low frequency content of the bass drums is generally attenuated while the amplitude of the bass notes is mostly maintained.

	RMS Amplitude		BSLoud		Log Frequency Centroid	
	UME NC	***UME HL***	***UME NC***	***UME HL***	***UME NC***	***UME HL***
n (samples)	597.00	597.00	299.00	299.00	598.00	598.00
Mean	-23.97 dB	-23.17 dB	16.38 dB	16.86 dB	1066.02 Hz	1076.33 Hz
ABS Diff.	0.81 dB		0.48 dB		10.31 Hz	
MAE	0.02		0.43		38.93	

Table 8.10: Quantification of properties of the track having NC and HL DRC.

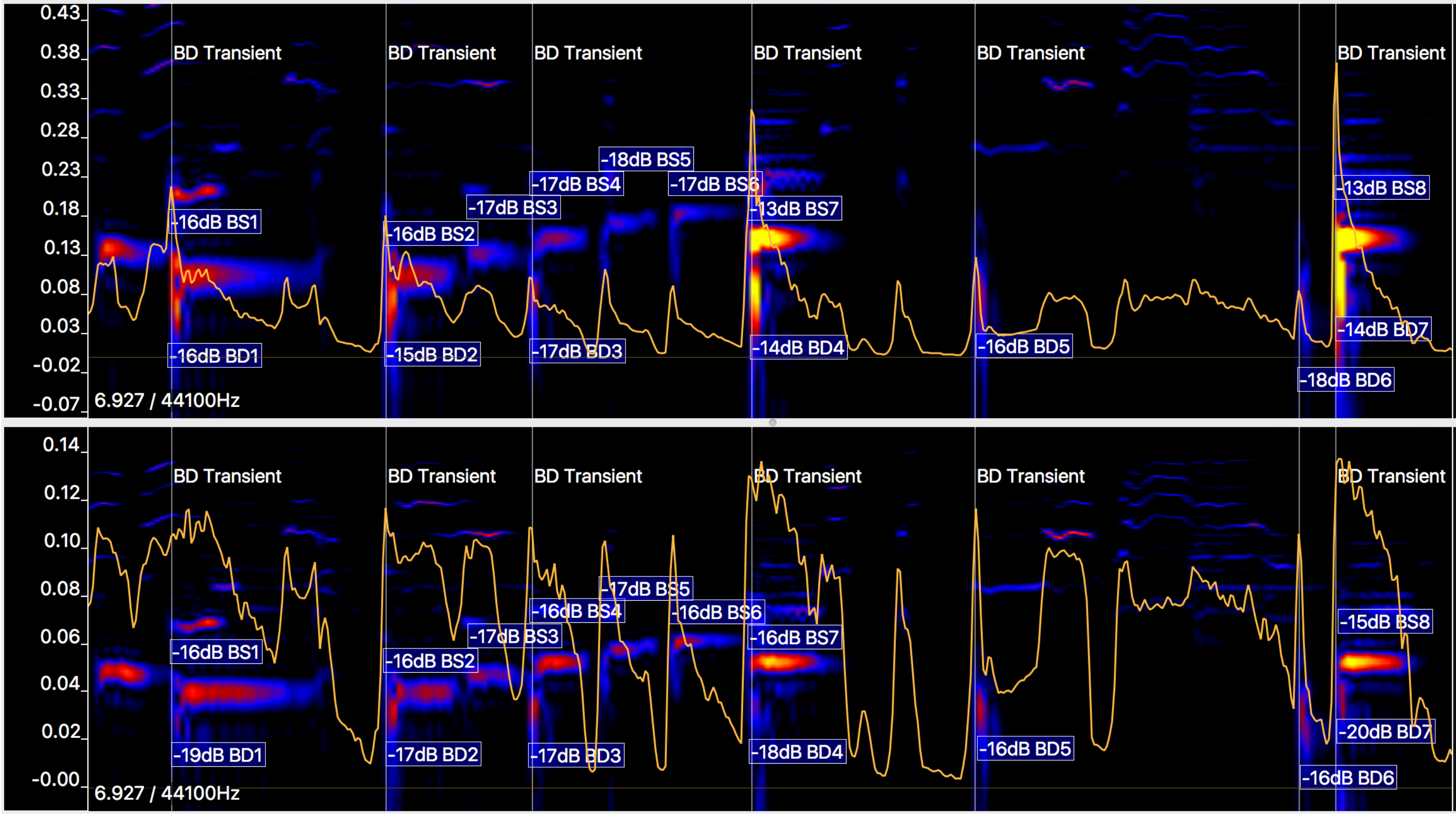


Figure 8.30: Spectrograph comparing the magnitude (dB) and frequency bins in the time domain, illustrating measured bass drum (BD) and bass guitar (BS) peak transient, followed a reference index for each; UME NC (top pane) and UME HL (bottom pane). Also included are plots demonstrating momentary RMS amplitudes (dB indicated on the left axis) indicated by the orange lines.

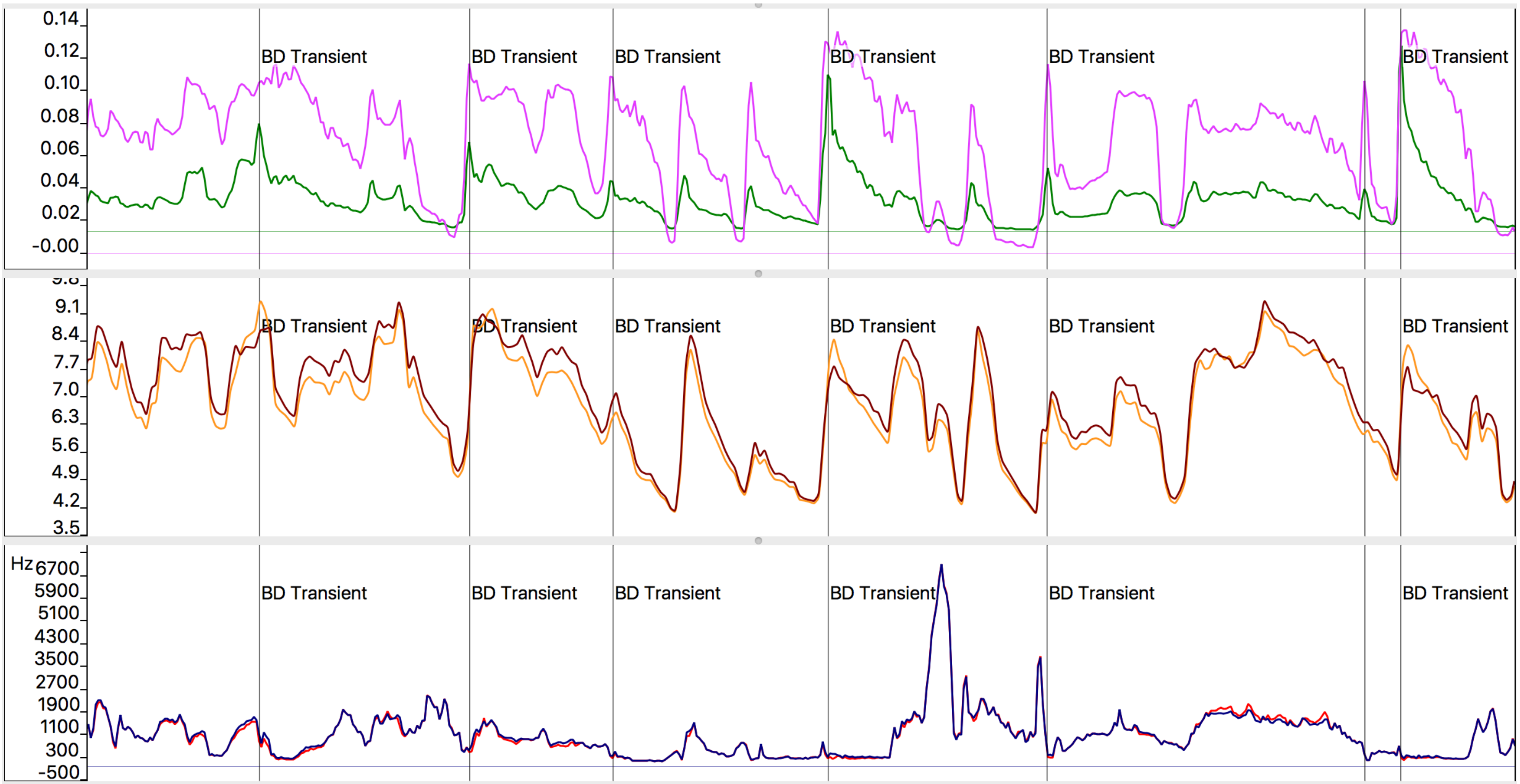


Figure 8.31: Transform analysis showing a like for like comparison of UME NC and HL including the reference points of each of the BD transients for reference; the top pane shows RMS amplitude (dB) vs. time (min:sec); the middle pane is BSLoud amplitude (dB) vs. time (min:sec); the bottom pane is LFC frequency (Hz) vs. time (min:sec).

B) Zone 2 Investigating harmonic corruption.

This section analyses a 22.5-second passage (from zone 2 Fig. 8.30) of a vocal signal (Fig. 8.32) using the SV transform: Harmonic Spectrum (Bullock, 2012). The transform calculates harmonic content by sampling only the top 10% (this is alterable on loading the transform) of a signal's peak amplitude, and only if calculated as harmonic. That is, the signal peaks are above the threshold of an integer multiple of the fundamental frequency. Figure 8.33 shows the resultant harmonic content of the NC vocal signal (top) and HL vocal signal (bottom).

Although visual inspection seems to reveal that there are differences between the two harmonic images, it is difficult to make conclusions without further quantification. The harmonic spectral transform data was exported to an Excel document for statistical comparisons using mean, ABS diff., and MAE.

The results exhibit noteworthy differences between harmonic content of the HL and NC stimuli (Table 8.11). The first observation is that the number of samples for HL is higher than NC, indicating increases of the peak signal content for HL i.e. more data within the top 10% of the waveform. Even though the differences indicated by the statistical measures are small, they demonstrate alterations to the harmonic content from applying DRC. The potential for this type of distortion was first demonstrated in Ward and Campbell (2010), which discussed audible tonal changes to instruments in re-mastered music; however, this is the first known quantification combined with statistical verification of this type of distortion.

Vocal Harmonic Content Comparison	***NC***	***HL***
n (samples)	62309.00	62925.00
Mean (dB)	-71.29	-67.99
ABS Diff. (dB)	3.31	
MAE relative to NC	0.000042	

Table 8.11: Statistical comparison of the harmonic configuration of a 22.6-second vocal passage, with and without DRC.

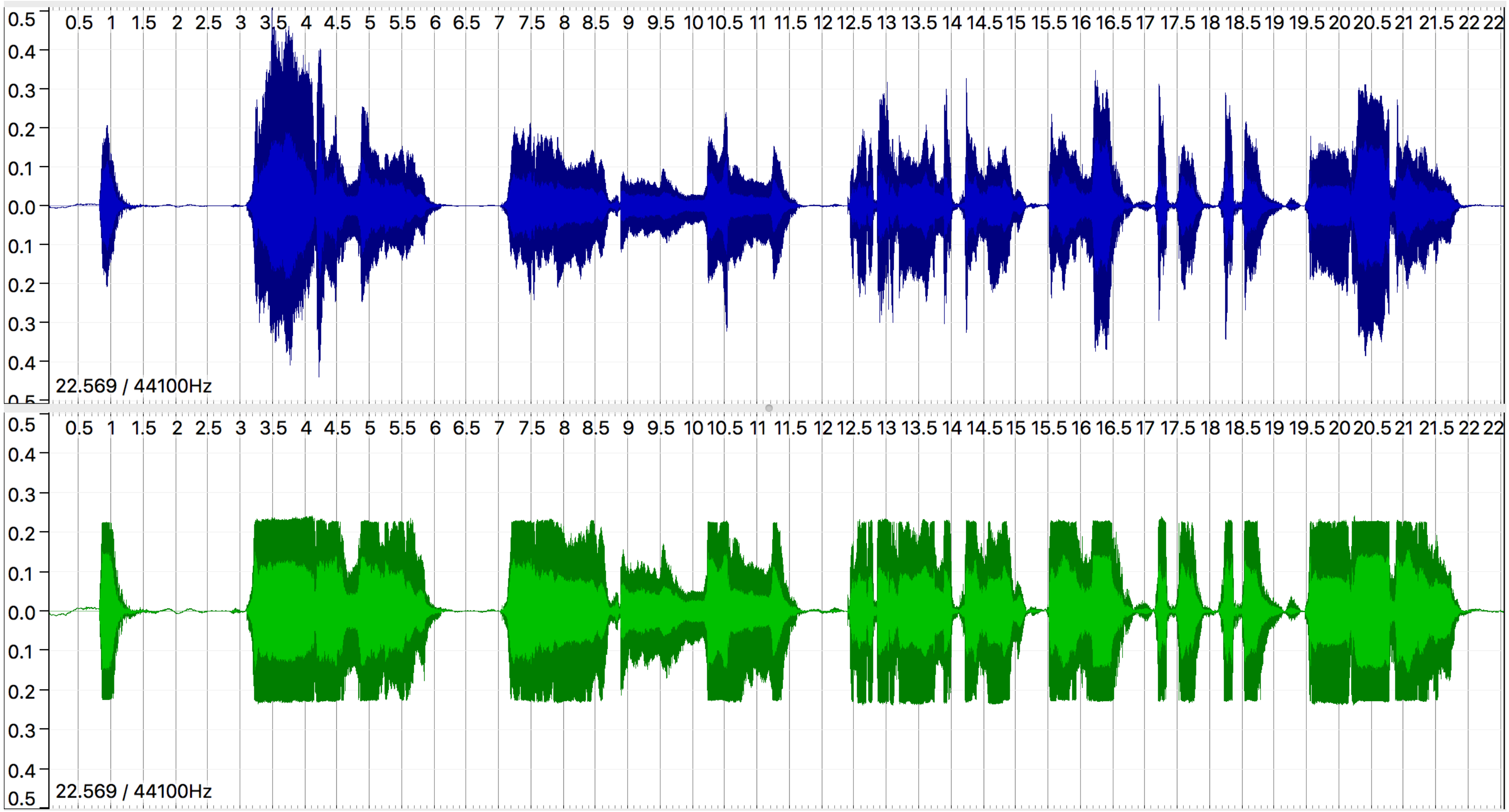


Figure 8.32: Waveforms of a 22.5-second vocal passage showing amplitude vs. time (s) NC (top) and HL (bottom).

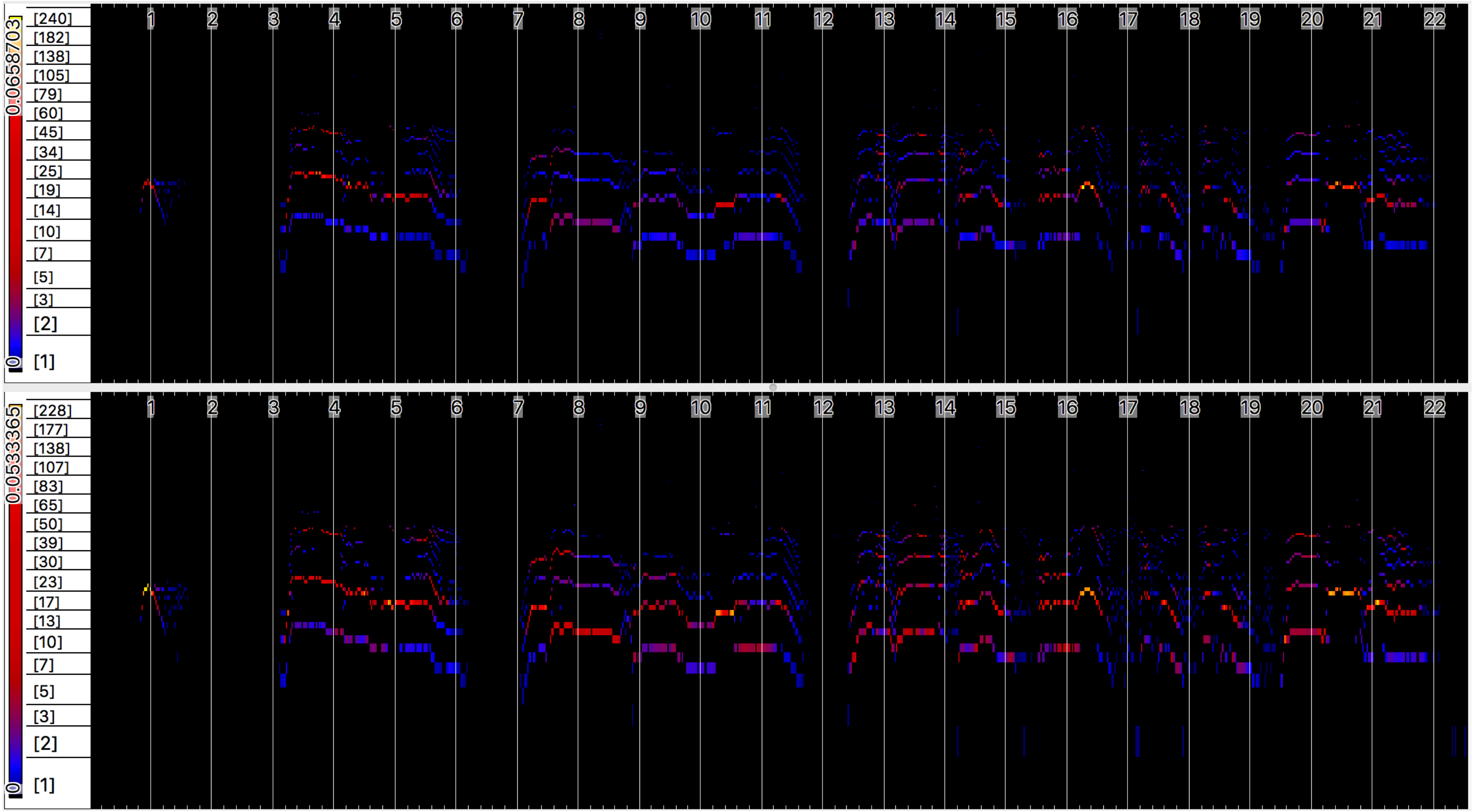


Figure 8.33: FFT showing the harmonic content of the 22.5-second vocal passage amplitude and frequency (Hz) vs. time, NC (top) and HL (bottom).

D) Zone 3: Instrument Interaction

Figure 8.34 displays the words corresponding to the vocal passage of zone 3 within the spectrogram, with lines extend downward demonstrating the relationship between the vocal, snare, and bass guitar notes. The bass signal shown, consists of 3 notes; sections 1 to 3 containing the 1st note (A2), sections 4 to 5 comprising the 2nd note (D♯3), and sections 6 to 10 including the final note (A2). NC (top) shows the bass signal as smooth sustained notes, while the corresponding HL bass signal appears to be made up of several notes. The effect is a product of DRC activation by the vocal line and deactivated when the amplitude of the vocal amplitude decreases (see section 2.1.8). The fast DRC release time allows the gain of the rest of the track to recover in amplitude, effectively altering the magnitude of the bass signal relative to the vocal signal (aka. pumping). DRC is also activated by the snare signal (indicated by the yellow arrows) like the vocal signal, producing similar consequences for the bass signal. Further quantification of this effect is in Test 2 below.

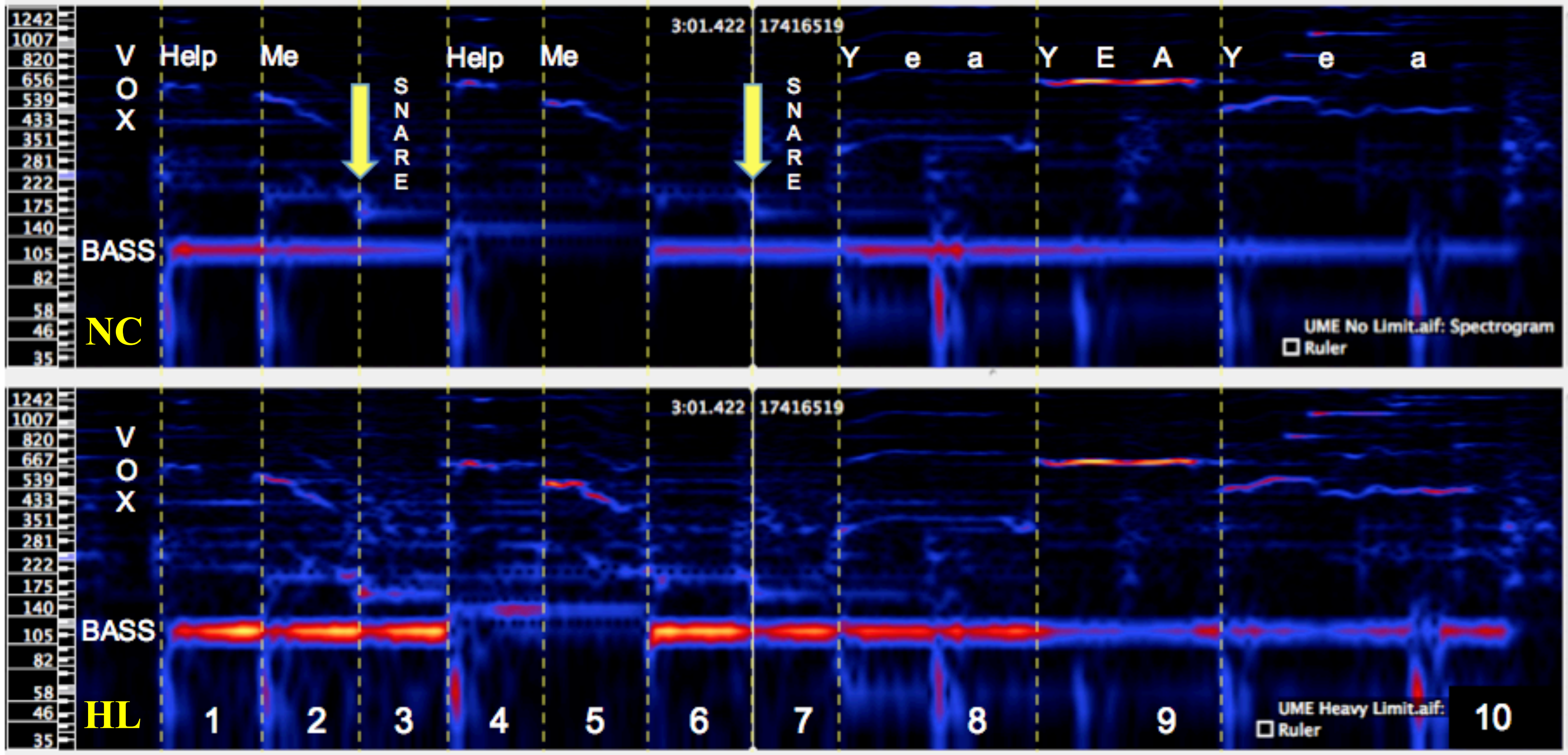


Figure 8.34: Spectrograph (amplitude and frequency vs. time) demonstrating pumping associated with the vocals, NC (top) and HL (bottom). The harmonic content of the vocal roughly aligns horizontally with the word 'VOX', typed vertically. The words of the vocal are annotated across the top, when they occur. The bass notes align with the word bass typed horizontally. The numbers 1-10 along the bottom of the figure are separating points of discussion.

8.5.2 Test 2: DRC Configuration

The following tests are based on the application of DRC at three different locations in the signal chain (Fig. 8.35): I) track DRC; II) subgroup DRC; and III) full-sum DRC. Visually, each of the waveforms demonstrates a highly dynamic appearance apart from full-sum DRC, with the subgroup configuration displaying the most controlled dynamics without the 'cut grass' appearance of full-sum DRC discussed earlier.

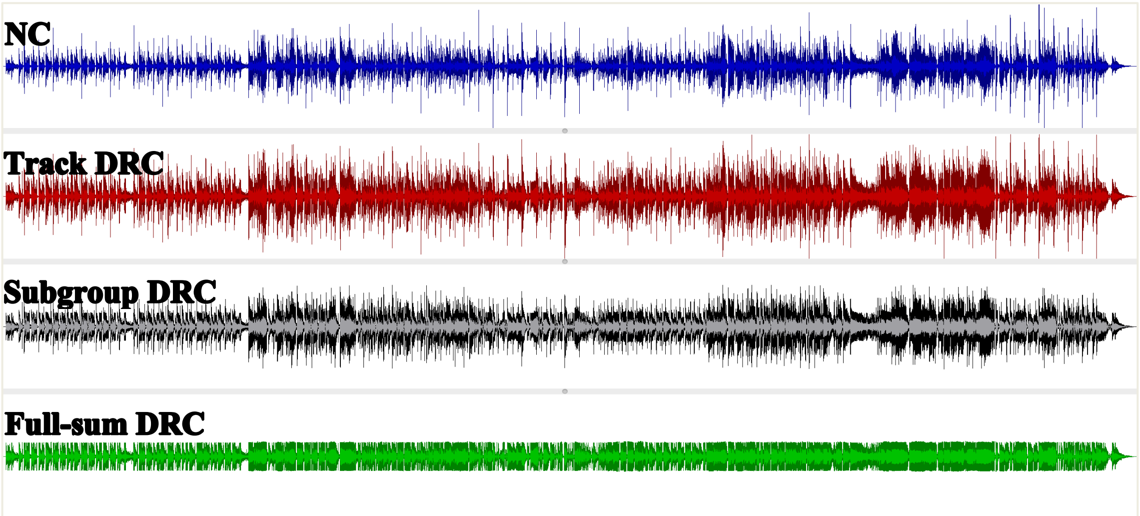


Figure 8.35: Waveforms (amplitude vs. time) of the DRC configurations compared: No DRC (top blue); followed by DRC applied at track level (red); followed by DRC applied to subgroup (black); followed by DRC applied to the full-sum (bottom green).

In Table 8.12 each configuration is tested against NC for loudness, RMS amplitude, and LFC, with comparisons of mean, ABS difference, and MAE. Figure 8.36 shows a seven-second excerpt of the RMS measurements plotted with reference to the corresponding NC spectrogram. Surprisingly, comparing the RMS amplitudes of the stimuli, track and subgroup DRC demonstrate more dynamic signals than NC while the full-sum configuration is similar to NC. The track plot demonstrates the detail of NC yet still exhibits control over the strongest peaks; the Subgroup plot is similar but at times appears to level out the amplitudes of the mix; the full-sum tends to moderate the whole of the signal (mostly) at the expense of transient sharpness. The track configuration has the highest mean RMS of -

19.35 dB followed by Subgroup (-19.86 dB) and full-sum (-20.56 dB). These statistical findings support the observation about track dynamics above.

	NC	***Track***	***Sub***	***Full***
	RMS			
n (samples)	18228	18228	18228	18228
Mean (dB)	-23.45	-19.35	-19.86	-20.56
ABS Diff. (dB)	Comparator	4.10	3.59	2.89
MAE relative to NC	Comparator	0.041	0.036	0.031
BSLoud				
n (samples)	9115	9115	9115	9115
Mean (dB)	16.24	17.32	17.29	17.21
ABS Diff. (dB)	Comparator	1.08	1.05	0.98
MAE relative to NC	Comparator	0.86	0.84	0.79
LFC				
n (samples)	18147	18145	18145	18144
Mean (Hz)	961.11	949.57	966.30	992.25
ABS Diff. (Hz)	Comparator	0.10	0.05	0.28
MAE relative to NC	Comparator	26.79	34.66	40.55

Table 8.12: Statistical analysis of the four DRC configurations, comparing BSLoud, RMS amplitude and LFC.

The first thing to note about the BSLoud (Table 8.12) is that full-sum DRC indeed controls on the dynamic loudness of a signal much more than the other two DRC configurations and therefore has potential for higher gain compensation as supported by the mean loudness of the full-sum being the lowest of the three DRC configurations. Figure 8.37 illustrates BSLoud at the micro-dynamics level, including a spectrogram (NC) for reference. There is an apparent loss of sharpness and detail for full-sum DRC when comparing the micro-dynamic loudness of all the configurations.

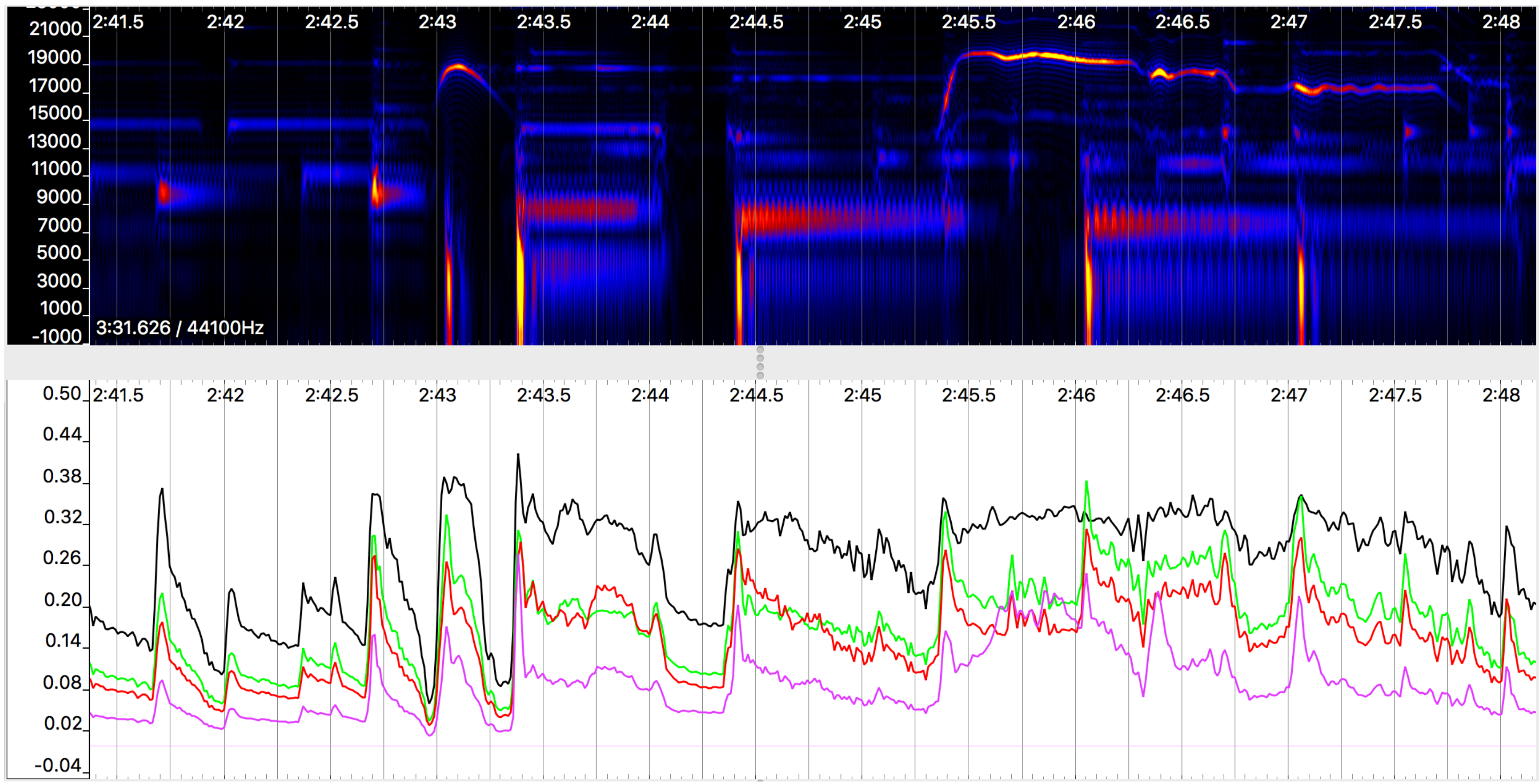


Figure 8.36: RMS amplitude over time of the four DRC configurations with a corresponding spectrogram showing frequency vs. time in the top window. The full-sum RMS plotted in black, subgroup is plotted in green, track level is plotted in red, and NC is plotted in magenta.

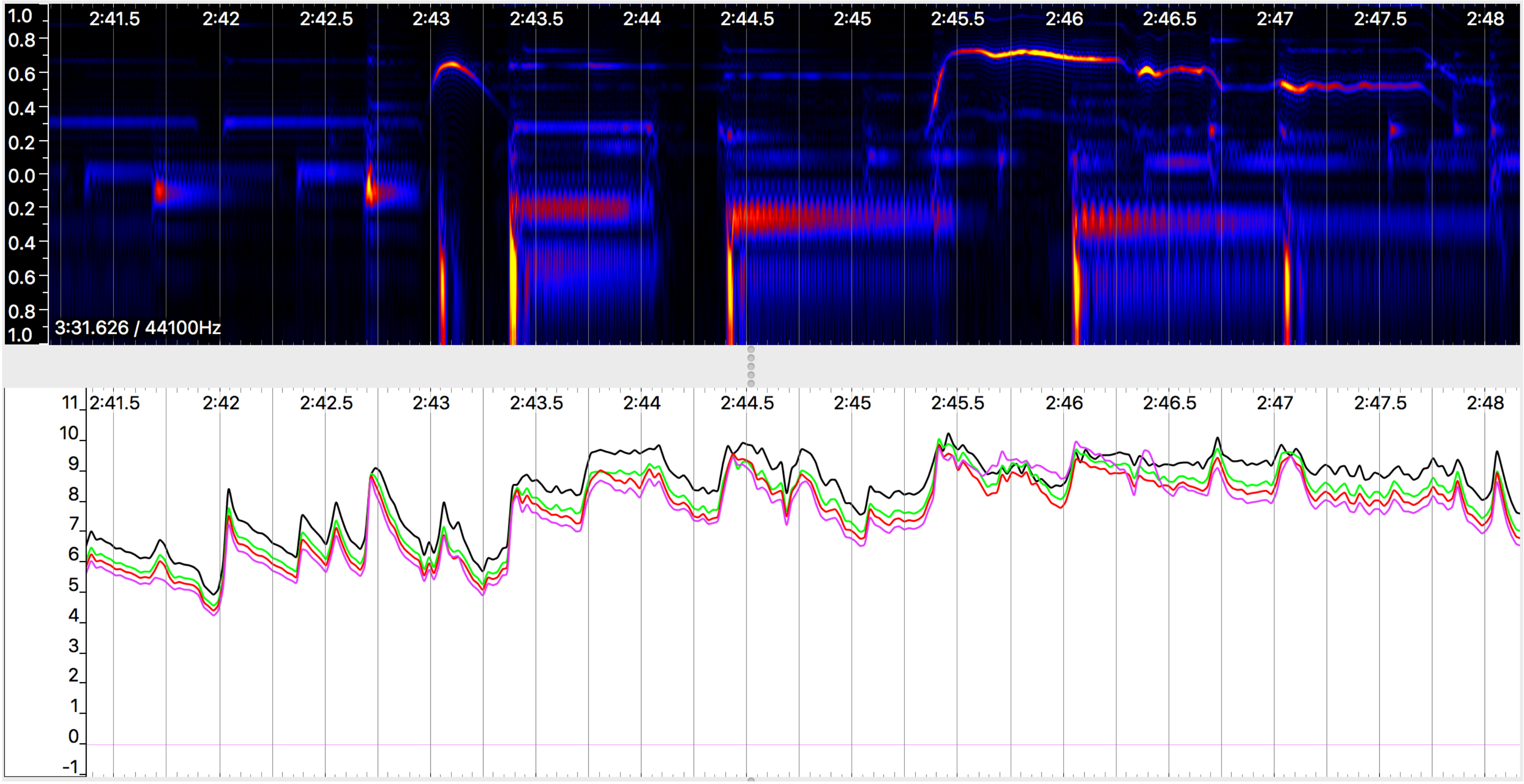


Figure 8.37: BSLoud (dB) against time of the four DRC configurations with a corresponding spectrogram showing frequency vs. time in the top window. The full-sum RMS plotted in black, subgroup is plotted in green, track level is plotted in red, and NC is plotted in magenta.

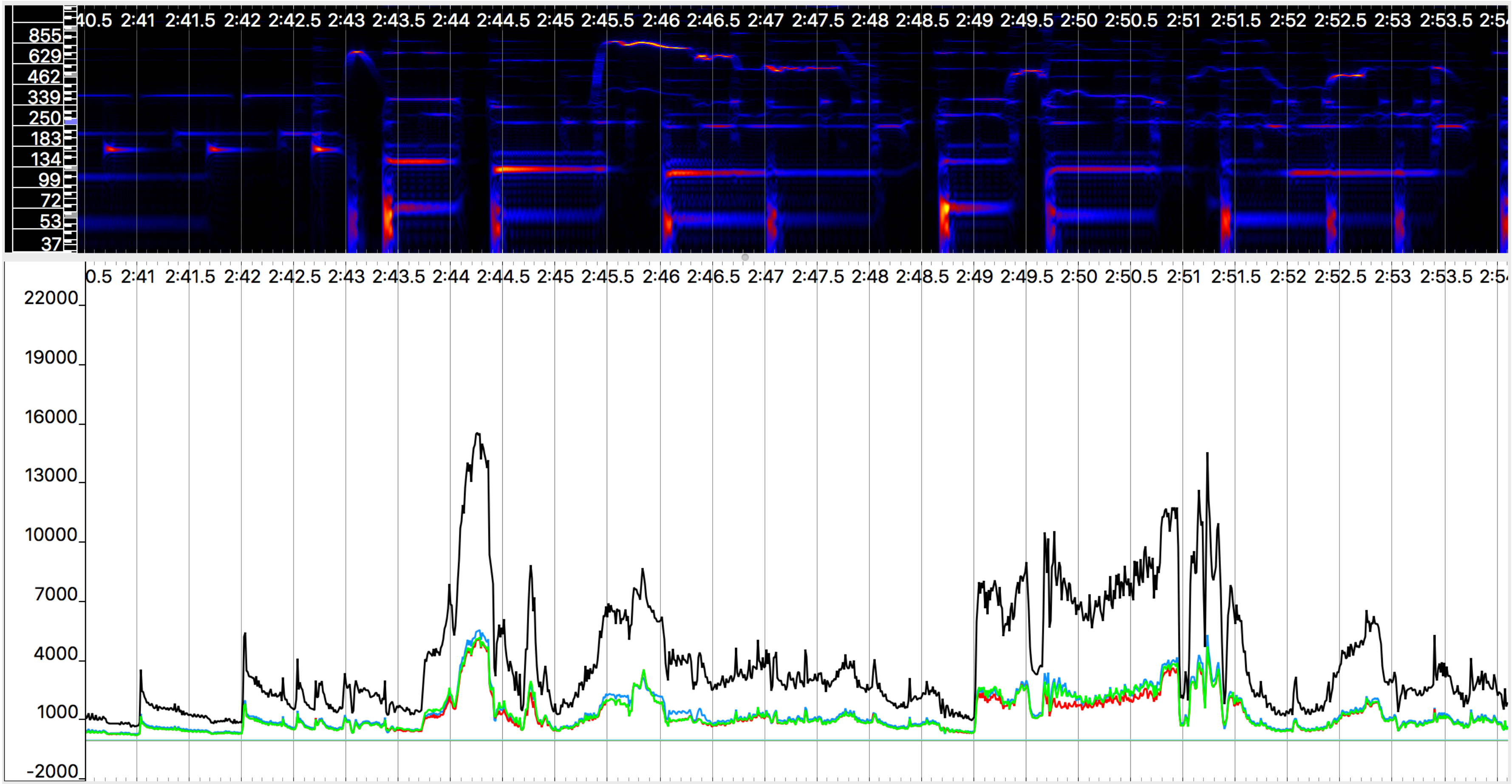
Figure 8.38 shows the LFC plots on top of a spectrogram, demonstrating distortion relative to each instrument. Interestingly, LFC gains occur between signals or during the decay phase of some of the instrument envelopes; most evident for full-sum DRC.

The Subgroup ABS diff. is only +5 Hz greater than the NC ABS diff., whereas track DRC has a 10 Hz upward shift and the full-sum has a 28 Hz upward shift in ABS diff. In other words, in this test, Subgroup DRC suffers the least distortion, followed by track, while full-sum DRC exhibits that most distortion.

Table 8.13 ranks each of the configurations (lowest MAE to highest), comparing DRC configurations based on loudness, RMS amplitude and LFC performance. These results seem to indicate that the increased distortion diminishes the loudness gain achieved by applying heavy DRC at the full-sum level. Put it in terms of ratio (loudness/noise) (where noise is assumed by the frequency shift indicated in LFC), DRC configurations: full-sum = 1/28 dB/Hz; Subgroup = 1/5 dB/Hz; and track = 1/10 dB/Hz. However, there are many ways to configure a Subgroup, and therefore these findings are only indicative.

Rank	Configuration	ABS Difference	MAE relative to NC
RMS Amplitude (ranked low to high MAE)			
1	Full-sum DRC	2.89 dB	0.03
2	Subgroup DRC	3.59 dB	0.03
3	Track DRC	4.10 dB	0.04
BSLoud (ranked low to high MAE)			
1	Full-sum DRC	0.98 dB	0.79
2	Subgroup DRC	1.05 dB	0.84
3	Track DRC	1.08 dB	0.86
Log Frequency Centroid (ranked low to high MAE)			
1	Track DRC	0.10 Hz	26.79
2	Subgroup DRC	0.05 Hz	34.66
3	Full-sum DRC	0.28 Hz	40.55

Table 8.13: Ranking of the four DRC configurations, comparing loudness, RMS amplitude and LFC.



ure 8.38 LFC (Hz) over time (min:sec) of the four DRC configurations with a corresponding spectrogram showing frequency (Hz) vs. time (min:sec) in the top window (this window is slightly larger to highlight the subtle differences between plots). The full-sum RMS plotted in black, subgroup is plotted in green, track level is plotted in red, and NC is plotted in magenta.

.

8.5.3 Test 3: DRC Configuration at Micro-dynamic Instrument Level

This test measures instruments interacting under the three DRC test configurations compared with NC. Figure 8.39 shows a 1s SV spectrogram excerpt of the Test 2 audio. Data amplitude measurements are taken at 5 μs intervals and entered into an Excel document for analysis. This part of the test focuses on the micro-dynamics of a one-second temporal window for five different instruments concurrently under DRC: vocal, guitar, bass, conga, and bass drum and how each instrument is affected by the different DRC configurations.

Figure 8.40 illustrates how DRC alters the spectral balance between the configurations; the degree of the effect is dependent on the DRC configuration as indicated by the alteration of the peak amplitudes (colour changes). Visual analysis shows alterations between the frequency cluster with the increased number of signals modulating under DRC. This can be most easily observed with the bass drum region being increased or decreased in magnitude in comparison with NC. The loss of contrast between NC and full-sum spectrograms is a function of IMD, as demonstrated in the MATLAB experiments.

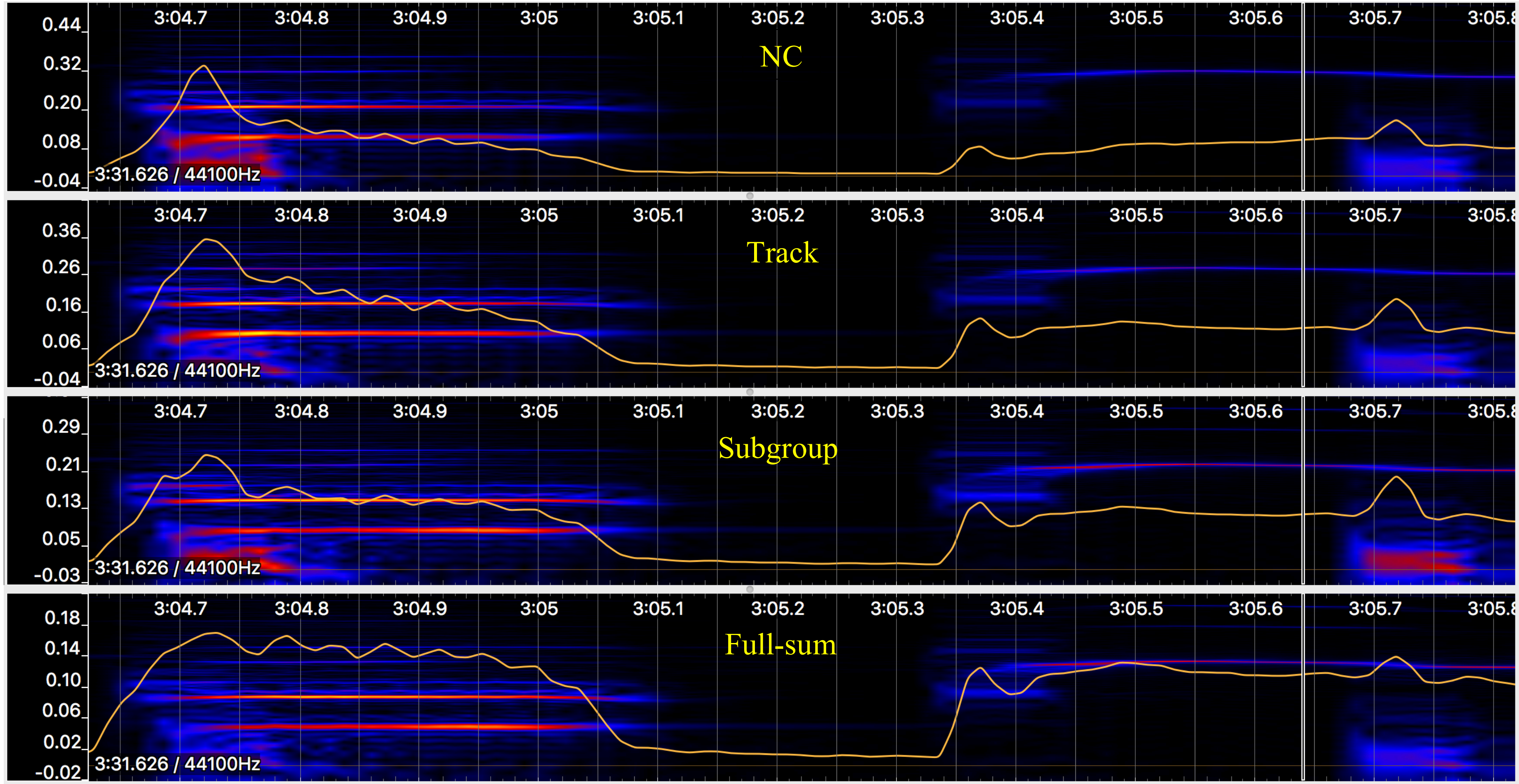


Figure 8.39: Spectrograms of the instrument magnitude (low is blue to medium yellow to high red vs. frequency on the y axis) over time (min:sec) of the mixed musical instrument signals under test showing the four DRC configurations with a plot of the RMS envelope (orange in dB).

Statistical analysis (Table 8.14) compares each instrument within each DRC configuration with the NC configuration using of mean, ABS difference, and MAE.

Fig. 8.39 charts the RMS of each instrument in each configuration. As one will notice, it is difficult to determine which configuration is optimal base on these measurements as the findings are different for each instrument.

NC					
Control	***Bass***	***Bass Drum***	***Conga***	***Guitar***	***Vocal***
n (samples)	22.00	22.00	22.00	22.00	22.00
Mean (dB)	23.15	25.08	26.67	22.19	24.25
Track DRC					
n (samples)	22.00	22.00	22.00	22.00	22.00
Mean (dB)	22.13	24.35	24.75	21.84	23.25
ABS Diff. from NC	1.02	0.73	1.92	0.34	1.00
MAE compared with NC	1.59	1.55	4.27	0.50	1.77
Subgroup DRC					
n (samples)	22.00	22.00	22.00	22.00	22.00
Mean (dB)	23.55	24.52	25.02	22.37	22.98
ABS Diff. from NC	0.40	0.57	1.65	0.18	1.27
MAE compared with NC	0.77	1.41	3.73	0.55	2.23
Full-sum DRC					
n (samples)	22.00	22.00	22.00	22.00	22.00
Mean (dB)	24.13	26.25	26.10	23.17	24.72
ABS Diff. from NC	0.99	1.17	0.57	0.99	0.47
MAE compared with NC	1.73	2.77	1.91	1.55	0.91

Table 8.14: Level measurements of 1 s excerpts of the signals under test.

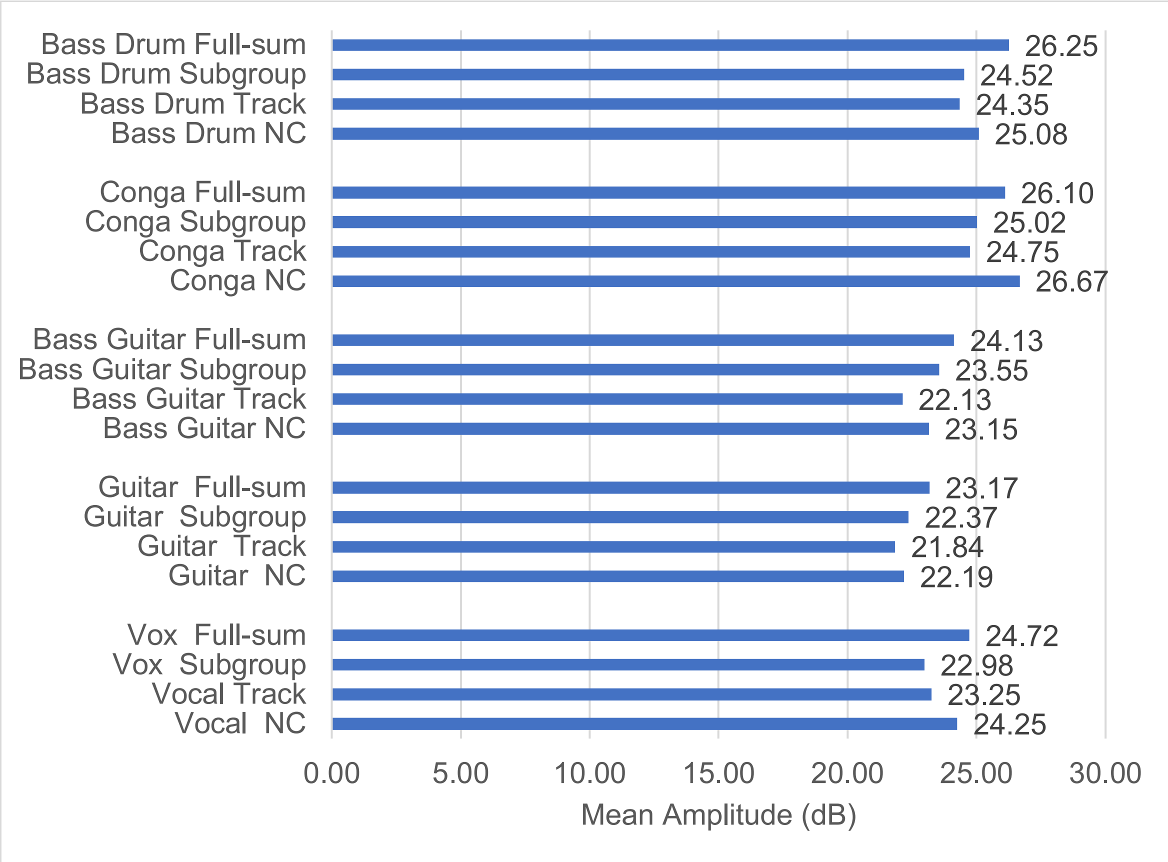


Figure 8.39: Mean amplitude measurements of the 1 s excerpts of the fundamental frequency of the signals under test.

Fig. 8.40 graphs measurements of the dominant frequency of each instrument of the four DRC configurations. The measured amplitudes do not indicate perceived loudness; instead, illustrate how conventional instruments might behave under the different DRC configurations compared with NC. Each plot shows the onset period of each instrument's envelope, followed by a rise in amplitude. Once at peak amplitude, each instrument has a short-sustained section, followed by a release section where the envelopes decline in amplitude.

It is apparent that the instrument envelopes change according to the type of DRC configuration. The harmonic frequencies related to each of the measured fundamentals will also be affected, as demonstrated in 8.5.1. Figure 8.41 shows the ABS diff. (i.e. the difference in dB) of each instrument from the NC control signal, demonstrating that the DRC configuration affects the signal envelope differently.

As discussed earlier, the onset phase (transient) of many percussive instruments is most important, as it allows the instrument to cut through the mix (punch) and establish the ‘beat’. It is therefore significant that the bass drum transient, and overall envelope, has such a considerable amplitude reduction under full-sum DRC, becoming the signal with the lowest amplitude. In fact, the onset of the bass guitar seems to suffer the same attenuation as the bass drum. These findings appear to be analogous to the results of the MATLAB sine wave experiments (section 8.3.2), which showed a reduction of the signal with the lowest frequency content modulating with other signals and deteriorating with additional signals.

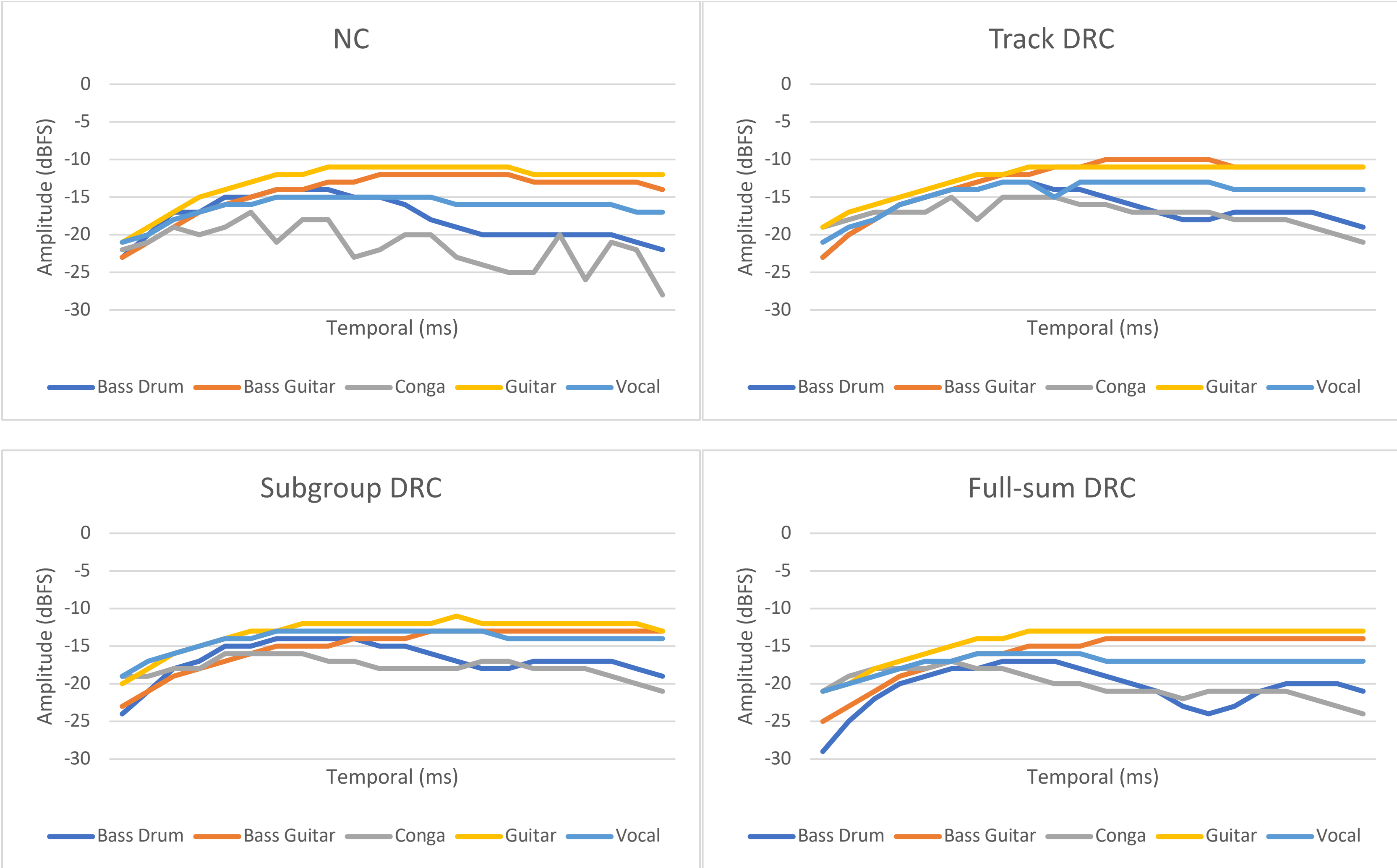


Figure 8.40: Level measurements of 1 s excerpts of the NC and DRC signals under test.

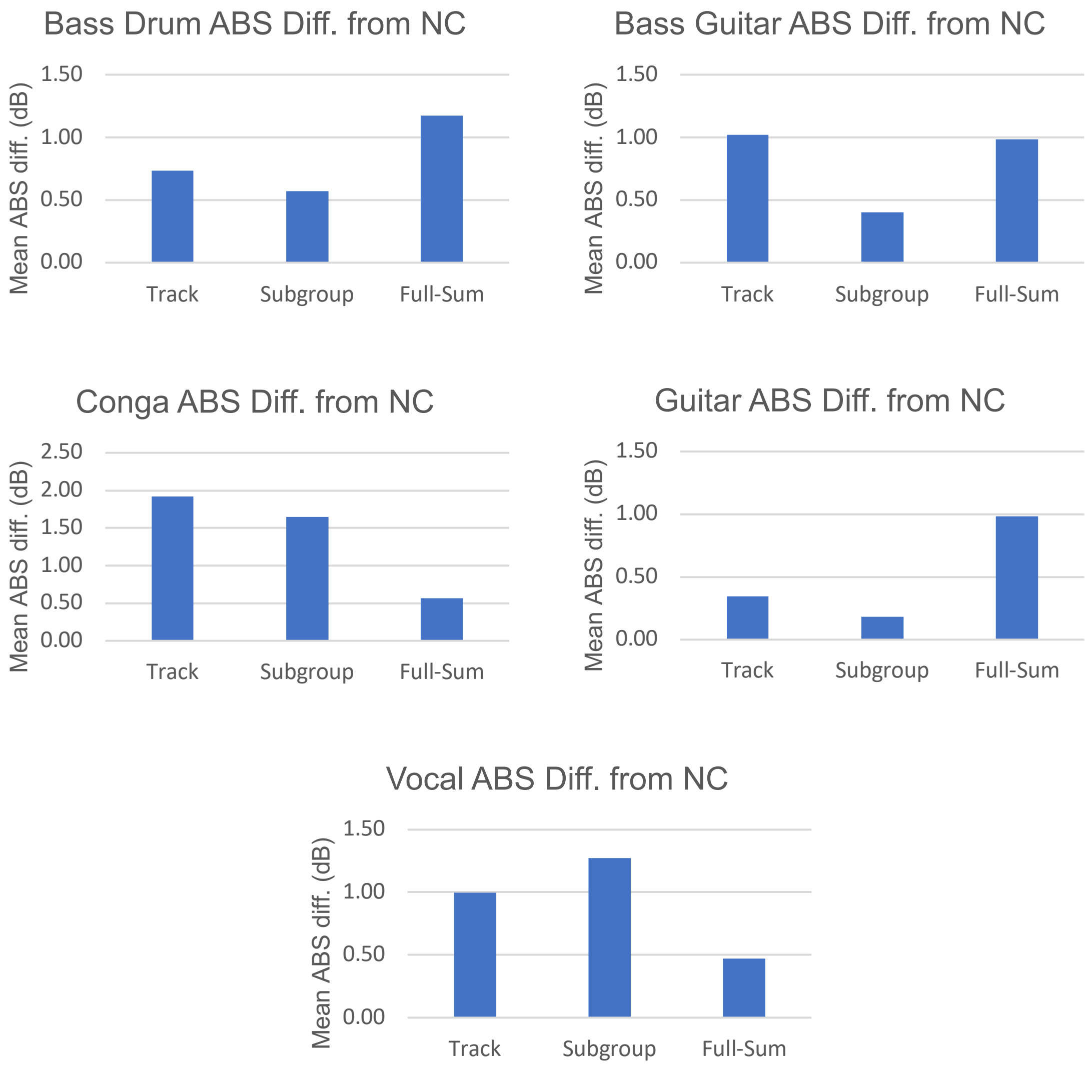


Figure 8.41: Graphs showing the ABS Differential for each instrument.

Fig. 8.42 illustrates the alterations further by displaying each instrument according to mean amplitude over the entire one-second window, relative the other instrument in each of the DRC configurations. This is further illustrated with the ranking of each instrument according to MAE in Table 8.15, which may indicate that different instruments differently according to the DRC configuration. However, the envelope of the conga is most notably altered by all the DRC configurations, as the NC envelope rises and falls rapidly in an oscillating fashion (a shown in the NC section of fig. 8.40). In every DRC configuration, its envelope bears little resemblance to NC, apart from the track DRC which seems to retain the first few oscillations.

The guitar seems to be the most stable under any DRC configuration as it demonstrates the most consistent ABS. diff. in fig. 8.42, with its most altered mean amplitude in the full-sum configuration.

Regarding controlling the dynamic levels of the overall mix, track and subgroup type DRC seems most consistent with NC. Both seem to reduce the dynamic variations without altering the balance of the instruments drastically. Contrarily, full-sum doesn't offer as much control of the overall dynamics of the signal and shows evidence of changing the balance between the instruments within the mix.

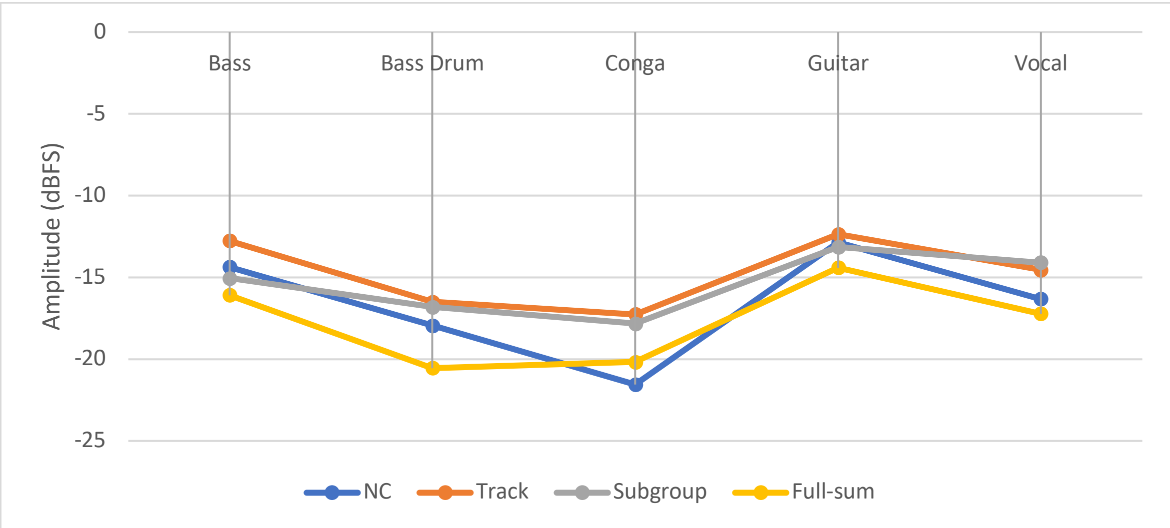


Figure 8.42: Relative instrument levels according to DRC configuration.

Rank	Configuration	ABS Difference	MAE
Bass (ranked low to high MAE)			
1	Subgroup	0.40	0.77
2	Track	1.02	1.59
3	Full-sum	0.99	1.73
Bass Drum (ranked low to high MAE)			
1	Subgroup	0.57	1.41
2	Track	0.73	1.55
3	Full-sum	1.17	2.77
Conga (ranked low to high MAE)			
1	Full-sum	0.57	1.91
2	Subgroup	1.65	3.73
3	Track	1.92	4.27
Guitar (ranked low to high MAE)			
1	Track	0.34	0.5
2	Subgroup	0.18	0.55
3	Full-sum	0.99	1.55
Vocal (ranked low to high MAE)			
1	Full-sum	0.47	0.91
2	Track	1.00	1.77
3	Subgroup	1.27	2.23

Table 8.15: Ranking each instrument by DRC configuration according to MAE of the 1 s excerpts of the signals under test.

8.6 Conclusions

This chapter explored the nonlinear effects of DRC first utilising sine waves and then with musical instrument signals. The sine wave tests included harmonically related and unrelated signals to evaluate the impact of DRC relative to musical tuning. The sine wave tests and instrument signal tests were performed first with signals having DRC applied after mixing (XMOD) followed by signals having DRC implemented before mixing (INDEP). This approach was intended to represent three different ways of applying DRC to music signals, consisting: track level DRC application; subgroup DRC application; and full-sum DRC application. In both test scenarios ‘extreme’ DRC settings were used to accentuate and illustrate the nonlinear effects of DRC, perhaps less observable for DRC (compression or limiting) applied conservatively and sympathetically to the signal. It must be stressed that this research does not advocate using DRC in such an extreme way (though there are numerous examples of such practice); clearly, there will also be more subtle ramifications in everyday use.

Use of DRC to increase the overall loudness of music programs is known to reduce dynamic variation. However, this research shows that DRC can be the source of rearrangement of instrument mixtures as well as the dynamic variations related to the rhythmic structure of musical signals. These rhythmic structures are often the defining aspect of a piece of music, so alteration may have the affect of changing the feeling and emotion of the track, quite possibly detrimentally. We have also shown that the harmonic structure of an instrument may be changed, potentially changing the tone of the instrument. For example, reducing the amplitudes of the low-frequency harmonic content of French horn many cause it to sound like a trumpet.

As the MATLAB experiments have demonstrated, IMD has the effect of adding sum and difference artefacts, harmonically unrelated to the fundamental frequencies of the notes, manifesting as noise. Increased magnitudes of DRC raise the proportion of IMD and shifts the LFC towards the higher frequencies. The degree of this shift is relative to the DRC configuration, full-sum much more so than track and subgroup DRC. We have shown that the ratio of loudness gain to noise is high for full-sum DRC, less for subgroup DRC and optimal for track based DRC.

LFC measurements also show that DRC can appreciably increase the amplitude of low-level sounds such as reverberation tails, delay effects, and even the natural release period of instrument envelopes.

This research has shown that decreasing the number of signals being compressed with each other while applying an equivalent amount of DRC reduces the IMD, thereby improving clarity and exerting more significant control of the instruments and mix dynamic variations. Also, reduction of the magnitude of DRC applied to a signal also reduces the extent of IMD artefacts.

Further research is required to understand better the low-frequency attenuation observed in the MATLAB and Pro Tools experiments. Future research should also incorporate testing the attack and release portions of DRC. Additionally, based on the final instrument level tests, further research should test the same and additional instruments under various conditions to see if their performance remains the same in the different DRC configurations.

Chapter 9:

General Discussion of Findings.

9.1 Introduction

This chapter discusses the combined findings of the experimental sections. Comparisons and explanations are made relating the four research questions. The chapter will include critical analysis of the combined results of each of the three experimental chapters.

9.2 Implications of the Experiments

This research described in this thesis implies that particular production strategies will reduce the negative effects of DRC on signals, verified by listener preference tests. The nonlinear characteristics of DRC, combined with the interaction of signals once summed, produce intermodulation distortion (IMD), which is unpleasant to hear. In a bid to reduce IMD, the point of application, along with the magnitude and type of DRC used in the mixing signal chain was tested, reducing the number of signals interacting. The different DRC configurations were also tested for fatigue and listener preference. Following is a review and discussion of the results organised according to the specific research questions the tests were intended to answer.

The first research question addressing the gap in knowledge posed the problem: "does listening to music subjected to heavy DRC lead to listening fatigue, evidenced by inferior performance in unrelated cognitive tasks?" This question was tested in chapter seven.

Anecdotally, listening to music with heavy DRC is fatiguing and may lead to eventual hearing damage (Rumsey, 2008; Jones, 2005; Levine, 2007; Vickers, 2010; Stone et al., 2009). Stone et al. (2009) showed that in specific scenarios augmented DRC on simultaneous speech signals impaired attention/concentration. However, research regarding fatigue from music subjected to heavy DRC being fatiguing is all but non-existent. The experiment discussed in chapter seven tested participants' sense of perceived elapsed time relative to listening to test stimuli with various DRC configurations.

The statistically significant findings suggest that all the test signals having DRC applied may be a source of fatigue. However, the results are not entirely clear, demonstrating and outcome opposite to what one would intuitively expect, and therefore are inconclusive at this time. Pooled together, all participants significantly overestimated the elapsed time, but none of the stimuli average estimations was substantially different from the true duration. Groupings combining magnitude, then type, followed by point were tested for significance, first by the mean overestimation, then by estimation error (how far the participant's estimations deviated from the actual time). These groupings indicated significant differences; however, this may only be an indication of universally inaccurate estimates collected rather than showing any specific/compelling DRC-setting related trends. The conclusion is the design of the experiment may be flawed; the primary limitation being the use of stimuli with too short a duration (20 seconds) to induce any fatigue symptoms, despite the recommendation by Zielinski, Rumsey, and Bech, (2008) that test stimuli longer than 30 seconds may introduce bias from the spatial and timbral characteristics of the test stimuli.

Results of these experiments indicate that time estimation means relative to magnitude and type DRC setting appear to decrease with increasing amounts of DRC applied. Heavy and limiting type DRC parameters, which compress signals the most had estimation means most like the actual test duration and estimated error. At the same time, moderate DRC and NC (least of all in all analyses) had estimation means furthest away from the actual test duration and estimated error.

Other researchers have shown that noise negatively impacts cognitive faculties (Rabbit, 1968; Sarampalis et al., 2009; Riley and McGregor, 2012), and it is therefore plausible that IMD related to the various DRC configurations used in these tests would influence the estimated duration of a musical excerpt. However, the SNR of these DRC stimuli may have been too high to induce fatigue-related symptoms. Rabbit (1968) used white noise correlated in intensity to the vocal passages that participants were to identify. In Sarampalis et al. (2009) the noise level (four people speaking simultaneously in the background) was maintained at 65dB SPL, and the stimuli were adjusted to either -2 or 2dB SNR. Riley and McGregor (2012) used white noise mixed with the to-be-learned words at +8 SNR (recommended classroom SNR is +15 SNR (Flexer, 2002). At moments there may be a 6dB or worse SNR in a heavily compressed musical signal (as measured in the test signals used in this thesis), but it is not a constant noise floor like the studies mentioned above and therefore perhaps negligible. SNR utilised by Rabbit (1968), Sarampalis et al. (2009), and Riley and McGregor (2012) are not realistic levels for even the least dynamic examples of contemporary music affected by DRC induced noise.

NC (no DRC) perhaps challenges explanations more so, as it deviates the furthest of all the stimuli from the 20-second benchmark. NC lacks the DRC processing of the other stimuli and therefore expected to exhibit a Mean Time Estimation (MTE) comparative to the baseline. Perhaps the findings of NC demonstrate that rather than finding evidence of fatigue, the temporal judgements may have been the result of distraction, enjoyment, or a combination of the three. Elapsed time can also correspond to pleasure – things seem to take longer when we're not enjoying them. For instance, Sackett et al. (2010) concluded that "… *people often neglect the duration of events when judging hedonic value.*" Therefore, we might expect music with the most IMD to be less agreeable, and to stretch time, relative to the most agreeable music. The participants only information before testing was that they would be listening to a piece of music, and may have assumed they would be subjectively judging the stimuli rather than judging elapsed time; therefore, perhaps the error bars are the real finding here; that as a whole, 'participants lost track of time, relative to the DRC configuration'.

Perhaps there is an unseen relevance concerning the exhibited direction of shorter and longer verbal estimations. Droit-Volet et al. (2004) postulate that over-estimations may indicate emotional arousal (happy, sad, angry), while Danckert and Allman (2005) suggest that underestimations may be suggestive of boredom. Noulhiane et al. (2007) found participants judged negative and emotional sounds as having longer durations than positive and emotionally neutral sounds. They concluded that the physiological responses produced by an emotional stimulus are "*the predominant aspect of the influence of emotions on time perception*". Over and underestimations found in our test perhaps may reflect the participants' internal clock is adapting according to environment events (Droit-Volet and Gil, 2009), such as distortions associated with nonlinear processing.

Stone et al. (2009) concluded that some listeners with less cognitive ability and below average hearing report an anecdotal sense of fatigue when listening to signals subjected to ‘heavy’ DRC, yet in these experiments, the findings seem to point toward conclusions other than fatigue. Perhaps fatigue is only a relevant in speech experiments and not significant with musical signals. As discussed in chapter three, auditory nerve (AN) transmissions are poor reproductions of auditory stimulus for frequency and amplitude. The prevailing theory proposes that the auditory periphery is explicitly adapted for species-typical vocal sounds rather than replicating all sounds uniformly (Bear, Connors, and Paradiso, 2007). In other words, perhaps DRC is fatiguing only where speech is concerned.

The anecdotal reports of heavy DRC music being fatiguing derives from music industry professionals predominantly (Rumsey, 2008; Jones, 2005; Levine, 2007; Vickers, 2010; Katz, 2007). Kirk (1956) found that habituation effects a participant’s perception of audio, and training can influence listening preferences either positively or negatively. Olive (2011) and Olive (2012) compared trained with untrained listeners and found that untrained listeners possibly could not hear differences in audio fidelity or were merely less focused during the tests for lack of interested in the music presented. Perhaps our participants (M_{age} = 22.35, SD_{age} = 5.57) previous listening experiences had a substantial impact on these findings? Perhaps listening to music with substantial levels of DRC is not fatiguing, maybe it is merely annoying to audio specialists, able to hear the artefacts imposed by heavy DRC usage.

The second research question addressing gap in knowledge ii posed the problem: which DRC configuration leads to the most significant rate of listener preference for a given piece of music.

Listener preference for loud music can be an influential factor under certain conditions (Packwood, 1974; Schubert, 2004a; Croghan, Arehart and Kates, 2012; Cullari and Semanchick, 1989; Neuhoff, McBeath, and Wanzie, 1999; Barrett and Hodges, 1995; Ronan et al., 2015). Cullari and Semanchick (1989) found a correlation between participants enjoyment of a piece of music and how loud they preferred to listen; Barrett and Hodges (1995) identified loudness preference relative to genre, with heavy metal and jazz preferred loudest; Ronan, Sazdov and Ward (2014) findings indicate that louder oration is more compelling. However, Croghan, Arehart and Kates (2012) found a decreased likelihood of stimuli preference and reduction in pleasantness ratings with the use of substantial DRC in both UNEQ and LEQ conditions. In contrast, De Man et al. (2015) found in the LEQ condition there was a distinct preference for more dynamic mixes. Nevertheless, studies have shown that the loudness of music has no significant correlation with commercial music success (Vickers, 2011; Viney, 2008).

For the reasons listed above, loudness was not a subject of investigation in this thesis. Instead, the focus was on developing a strategy to minimise the effects of nonlinear distortion through manipulation of the DRC configuration, most specifically the point of DRC application within the mix chain. There are several research projects having tested DRC configurations and participants arrangement preferences (Maddams, Finn and Reiss, 2012; Giannoulis, Massberg and Reiss, 2012; De Man and Reiss, 2013; Pestana and Reiss, 2014; De Man et al., 2014; De Man et al., 2015; Ma et al. 2015). The studies determined reasons for subgrouping, along with the exploration of various instrument subgroup configurations (Ronan, Gunes and Reiss, 2017). It is worth noting that the approach of studies to date have based their DRC arrangements on current and historical practice, rather than developing best practice for minimising the effect of DRC. There are currently no known published studies (other than Campbell, Paterson, van der Linde, 2017) that examine DRC configurations designed to reduce IMD, nor testing such configurations for listener preference.

The first experiment in this thesis asked untrained listeners to rate musical stimuli subjected to different DRC configurations preferentially. The experiment design was to ascertain the impact, independently and in combination, of DRC magnitude (moderate or heavy), type (compression or limiting), and the point of DRC application (track, subgroup or full-sum). The findings indicate preferences for DRC applied to fewer signals modulating together (track and less so, subgroup), as opposed to compressing multiple summed signals as is a traditional method for music production (as well as combinations of the three methods), and the use of compression rather than limiting.

These findings align with those of Stone et al. (2009) who found that applying DRC after signal summation (XMOD) (subgrouping or full-sum) the combined effects of magnitude and point, adversly affected task performance. They also found that applying DRC before summation (INDEP) produced less distortion relative to XMOD.

Maddams, Finn and Reiss, (2012) examined altering DRC combinations of ratio (type) and threshold (magnitude). They consistently found participant preference for 'threshold–light' DRC magnitude, stipulating that light DRC application was generally the most subjectively appropriate. They also found that participants consistently rated the 'threshold-heavy' condition poorly. Listener preference in these experiments show stimulus groups selected less often than expected by chance as L (limiting type); HL (heavy magnitude, limiting type); HF (heavy magnitude, full-sum point); LF (limiting type, full-sum point); F (full-sum point) and HLF (heavy magnitude, limiting type, full-sum point). The apparent stimuli relationship shows a negative preference for limiting when using a heavy magnitude, applied at the full-sum point in the mix chain.

Notwithstanding, the findings in this thesis related to the NC configurations is somewhat inexplicable; NC was consistently chosen as predicted, i.e. no preference, positive or negative. Perhaps a partial explanation is that participants generally prefer DRC stimuli over NC as was found in Maddams, Finn and Reiss (2012) and Ma et al. (2015). The Mean age of our participants was 22.35, (SD_{age} = 5.57) and it is probable that they have predominantly been exposed to music with substantial DRC and thus predisposed to a preference for DRC. Kirk (1956) found that habituation affects the participant's perception of audio, and training can influence listening preferences either positively or negatively. Findings of Olive (2011) and Olive (2012) also showed that a plausible relationship exists between expertise and general appreciation for audio fidelity. Had participant been trained to hear nonlinear artefacts, as recommended by Reiss (2016), results regarding NC stimuli. Still, the use of untrained listeners seems more relative to contemporary practice of DRC usage in the production of popular music, though here we only tested one genre; Ma et al. (2015) suggest that listeners may prefer NC, depending on their DRC preference relative to the musical style.

The third research question, addressing the gap in knowledge iii, posed the problem: "what is the quantitative impact of the nonlinear properties of DRC on real music signals?" Existing studies suggest that the use of DRC for loudness maximisation compromises audio fidelity (Aarseth, 2012b; Croghan, Arehart and Kates, 2012; Essling, Koenen and Peukert, 2014; Wendl and Lee, 2014; Deruty and Tardieu, 2014; Kirchberger and Russo, 2016; Vickers, 2010; Viney, 2008; Pestana and Reiss, 2014; Gorlow and Reiss, 2013). Stone et al. (2009) found that DRC applied to the spoken word affects the temporal contrast (amplitude variation and content over time), the spectral contrast (amplitude variation and content of the frequency spectrum) of the signal and is dependent on the number of signals having DRC applied simultaneously, onset and release speed, and DRC magnitude. Whether these observations generalise to musical signals was unclear, leading to the third set of experiments exploring the nonlinear properties of DRC on simple sine wave configurations followed by music stimuli.

The effects of DRC on musical signals are rarely quantified beyond amplitude measures, i.e. dynamic range, crest factor, loudness range etc. Investigations in production of this thesis are unique in the methods of quantifying the effects of DRC on musical signals (Campbell, Toulson, and Paterson, 2010; Ward, and Campbell, 2010; Campbell, 2012; Campbell, Paterson, and Toulson, 2014; Campbell, 2014).

The analysis performed in this thesis uses extreme DRC (yet comparable to some contemporary music productions) to understand the associated nonlinear behaviour of DRC by emphasising the distortions present in all DRC signals (compression or limiting), perhaps less apparent for DRC applied conservatively and sympathetic to the signal envelope. The sine wave and music tests included signals having DRC applied after mixing (XMOD) followed by DRC application to signals before summation (INDEP). The sine wave tests represent three different ways of applying DRC to music signals, consisting: track level DRC application; subgroup DRC application; and full-sum DRC application. Sine wave analysis provided valuable insight and guidance for review of the music signals, supported by a statistical analysis utilising various measures to quantify visual observations.

One of the most compelling observations demonstrated by the sine wave tests and further observed in the music tests is the attenuation of low-frequency components or other rearrangements of micro-dynamic properties of modulating signals. The sine tests showed that XMOD HRL stimuli are affected more than HUR stimuli, and INDEP stimuli were not affected to the same degree, regardless of their tonal relationship. The same effect is observed with the musical signals, though not as predictably, and therefore requires further future experimentation. Factors that may influence the subsequent attenuations or other micro-dynamic rearrangements observed in the musical signals include the following: the harmonic relationship of the notes; temporal alignment (e.g. the bass drum transient may occur at any infinitesimal time ranging from before to following the other simultaneous signals); the velocity of each instruments signal envelope; and any combination of each aspect. However, this research demonstrates that avoiding this anomaly is achieved with the minimisation of intermodulation signals by altering the point of DRC application where possible.

Another critical observation relating the sine tests to the musical tests is the shift in the log frequency centroid (LFC) relative to DRC and the various DRC configurations. Application of DRC to a single sine wave introduces an LFC shift, evident by the addition of corresponding frequency components (see Fig. 8.11). The same effect is compounded with

XMOD HRL signals and worsened with HUR signals; when DRC utilisation is before mixing (INDEP), the LFC is minimal and comparable with the LFC of a single sine under DRC. Subsequently, the frequency shift evident with the sine wave tests are also apparent with musical signals. Interestingly, DRC application at the track level shifts the LFC toward the low-frequencies, which may be the result of the low-frequency envelope of the bass drum being intensified by DRC (see Fig. 8.35). The difference between the LFC of the DRC configurations is most evident in Fig. 8.37, which demonstrates the improvement achievable from altering the location of DRC, not so apparent from the statistical evidence. Clearly, the LFC of the track and subgroup configurations are relatively similar to NC; however, the full-sum configuration differs the most from NC, illustrating the effect of DRC related IMD perfectly. It is worth noting that although DRC applied at the track level seems to generally outperform the other configurations, the subgroup configuration may lead to a more controlled signal while maintaining reduced distortions observed by quantifying LFC. Additionally, observations made on the individual instrument envelopes suggest that micro-dynamic distortions are evident in all the DRC configurations.

The final research question addressing the gap in knowledge iv posed the problem: “can DRC configuration manipulation minimise fatigue, maximise listener preferences, and quantitative signal quality?”

This research has not identified the optimal DRC configuration for minimising fatigue due to the lack of clarity in the results for the fatigue tests. However, DRC configuration can maximise listener preference, and preserve signal quality, partially answering the final research question: the DRC application point in the mix chain that listeners prefer for music applied DRC to fewer signals simultaneously (i.e., to tracks before grouping/summation). The type of DRC preferred by listeners was compression over limiting. Listener magnitude preference was for moderate DRC over heavy and no DRC.

Therefore, the optimal configuration for listener preference is moderate compression applied at the track location. This research has shown that decreasing the number of signals mixed together while using an equivalent amount of DRC reduces the IMD, thereby improving clarity, and exerting significant control of the instrumentation and mix dynamic variations. Also, reduction of the magnitude of DRC applied to a signal reduces the creation of IMD artefacts. Negative preferences may be explained by the combination of sine wave analysis applied to musical signals, confirming that IMD (associated with XMOD configurations) adds sum and difference distortion artefacts, harmonically unrelated to the fundamental components of the signal. These nonlinear distortions manifest as noise, observed in this thesis using LFC quantification, and described by listeners as ‘harshness’ or ‘roughness’ (Moore and Tan, 2003; Tan et al., 2003).

Increased magnitudes of DRC raise the proportion of IMD and shifts the frequency content of signals (sine waves and music) towards the higher frequencies as quantified using LFC. The degree of this shift is relative to the DRC configuration, full-sum much more so than track and subgroup DRC. In fact, the ratio of loudness gain to noise is greatest for full-sum DRC, less for subgroup DRC and track based DRC configurations: full-sum = 1/28 dB/Hz; subgroup = 1/5 dB/Hz; and track = 1/10 dB/Hz.

9.3 Limitations and Future Work

The use of DRC to increase the overall loudness of music programs is known to reduce dynamic variation. These research findings demonstrate that DRC can effectively rearrange relative signal levels between instruments. The rearrangement is also dependent on other features such as DRC point and type, amplitude, temporal alignment and what instrument(s) DRC is acting upon. Instrument type (e.g., drums, vocals and other instruments) determined their organisation into subgroups; this research has shown that there may be optimal configurations according to the instrument properties to optimise DRC usage which requires further testing.

DRC can alter the dynamic variations related to the rhythmic structure of musical signals, often the defining aspect of a piece of music. DRC can also rearrange the harmonic structure of each instrument within a mix, changing the arrangements tonality. The above experiments have shown these alterations are the product of IMD. Further research is required using a wider variety of conditions to understand the phenomenon, specifically relative to remastering practice. As shown in Ward and Campbell (2010), this may have a detrimental effect on historical recordings during restoration, archiving and republication.

Further research is required to understand better the low-frequency attenuation observed in the XMOD experiments. Although experiments in this thesis were able to observer and predict the phenomenon for the instrument signals, based on the findings of the sine tests, there is no clear explanation for why it happens.

In this thesis, the DRC attack and release parameters were set to function in a general way, rather than being specially optimised for each stimulus configuration, which may have influenced listener preferences. However, it would be difficult to obtain quantitative and comparable results had the DRC parameters been subjectively adjusted according to the variability of the individual signal envelopes. Future research should also incorporate testing the attack and release portions of DRC. Future testing should also extend to include examining the various DRC parameters using stereo signals and analysis of the stereo spectrum.

In relation to the fatigue tests undertaken in chapter eight, there may be an unseen relevance concerning the direction of MTE deviations. Droit-Volet, Brunot, and Niedenthal (2004) and Droit-Volet, Tourret, and Wearden, (2004) suggest that over estimations may indicate emotional arousal (happy, sad, angry), while Danckert and Allman (2005) suggest under estimations may express boredom. Noulhiane et al. (2007) found negative sounds, and low-arousal stimuli judged longer than positive and high-arousal stimuli.

A single music excerpt provided listening test stimuli for chapters six and seven experiments, potentially restricting the degree to which we can generalise our preference and MTE findings to other genres. Future research should consider lengthening the MTE test so that participants can estimate the times for each DRC configuration for the various musical styles, with varying lengths, however, this would likely be an extensive process. Furthermore, additional research should verify whether these results generalise to different music genres (i.e., to a broader range and combinations of timbres), durations, instrumentations, and musical structures, and to examine the impact of DRC ratio, attack, and release parameters on listener preferences more exhaustively.

Chapter 10:

Conclusions

10.1 Conclusions

The problem addressed in this research was to define how DRC affects the psychoacoustic quality (the basic signal makeup of musical audio that influences our enjoyment of listening to music) and loudness (which may influence listening fatigue) of commercial music, and influence change in practice, drive audio quality education and inform best practice.

This thesis tested the effect of DRC on musical signals. The experimental approach analysed both simple and musical signals both numerically and in listening tests with human participants. DRC configurations designed to minimise the effects of nonlinear signal processing were tested against listener preferences; the potential impact of musical signals subject to DRC on listener fatigue was examined, and the propensity for DRC to degrade signal quality was quantified. New DRC configuration recommendations were formulated that are intended to produce the least IMD, and maximise listener preference.

Use of DRC to increase the overall loudness of music programs is known to reduce dynamic variation. However, this research has shown that DRC can change dynamic organisation related to the rhythmic structure of musical signals. This research has shown that decreasing the number of signals mixed with each other while applying an equivalent amount of DRC can reduce IMD and thereby improve clarity. Also, reducing the magnitude of DRC applied to a signal reduces the creation of IMD artefacts. The combination of strategies listed, significantly enhances signal clarity by lowering the DRC modifications of the original signal. As a result, listener enjoyment increases dramatically, as demonstrated here using listening tests.

One previously unknown consequence of the nonlinear properties of DRC revealed was that the harmonic content of a vocal passage was significantly altered by DRC. The modifications manifest with the alteration of the harmonic content of the recorded voice, recognisable as tonal alterations to the signal.

As the experiments in chapter eight demonstrated utilising simple sine wave configurations, IMD fills in the frequency bands around the original frequencies, dependent on the DRC configuration. Increasing DRC magnitudes equate to raising the amplitude of the IMD. These experiments have shown that IMD affects temporal clarity. DRC has the effect of increasing the amplitude of quiet sounds such as reverberation tails, delay effects, and even the natural release period of instrument envelopes.

Listening test experiments attempted to confirm the anecdotal theory that heavy DRC induces listening fatigue. Unfortunately, the test results proved to be inconclusive, likely a consequence of limitations in the experimental design.

10.2 Contributions to Knowledge

a) Statistically significant listening test results demonstrate that the manipulation of the DRC process chain leads to the most preferable music. Specifically, that DRC preference is for application to fewer signals simultaneously (i.e. track DRC is preferred over subgroup DRC, and both are preferred over full summation DRC). (addressing gap in knowledge *ii*).

b) The nonlinear properties of DRC applied to real music signals are similar to those described in Stone et al. (2009) for the spoken word. Specifically, analyses demonstrate that DRC affects the temporal and spectral contrast of the compressed signal, dependent on the number of signals having DRC applied simultaneously and DRC magnitude (note: DRC onset and release speed were not tested). (addressing gap in knowledge *iii*).

c) DRC configuration for maximising listener preference and preserving signal quality may be achieved by manipulating the point in the mix chain where DRC is applied, *viz*., it is preferable to apply DRC to fewer signal simultaneously, use compression rather than limiting, and to apply light to moderate DRC over heavy or no DRC. (part-addressing gap in knowledge *iv*)

References

ATSC, A. 85, 2009. Techniques for Establishing and Maintaining Audio Loudness for Digital Television, Washington, DC: Advanced Television Systems Committee.

Aarseth, B., 2012. DAB Norway–Implementation of Loudness Normalization. EBU Technical Review. Available at: https://tech. ebu. ch/docs/techreview/trev_2012-Q1_dab-loudness_aarseth. pdf.

allmusic.com, 2016. Bob Katz Credits. [on-line] Available at: <http://www.allmusic.com/artist/bob-katz-mn0000054492/credits>.

allmusic.com, 2016. Bob Ludwig Credits. [on-line] Available at: <http://www.allmusic.com/artist/bob-ludwig-mn0000067819/credits>.

allmusic.com, 2016. Joe Meek Credits. [on-line] Available at: <http://www.allmusic.com/artist/joe-meek-mn0000209088/biography>.

Arehart, K.H., Kates, J.M. and Anderson, M.C., 2011. Effects of noise, nonlinear processing, and linear filtering on perceived music quality. International Journal of Audiology, 50(3), pp.177–190.

Arfib, D., 1978. Digital synthesis of complex spectra by means of multiplication of non linear distorted sine waves. Audio Engineering Society Convention 59. Montreal, QC, Canada, 15-17 July 2015. New York: Audio Engineering Society.

Baddeley, A., 1998. The central executive: A concept and some misconceptions. Journal of the International Neuropsychological Society, 4(5), pp.523-526. Cambridge: Cambridge University Press

Baddeley, A., 2000. The episodic buffer: a new component of working memory? Trends in Cognitive Sciences, 4(11), pp.417–423.

Baddeley, A.D. & Hitch, G., 1974. Working Memory. Psychology of Learning and Motivation, pp.47–89.

Ballou, G., 2013. Handbook for sound engineers. 4th ed. Abingdon, England, UK: Taylor & Francis.

Barrett, D. and Hodges, D., 1995. Music loudness preferences of middle school and college students. Texas Music Education Research, pp.1-6. Austin, USA: Texas Music Educators Association.

Bear, M.F., Connors, B.W. and Paradiso, M.A., ed., 2007. Neuroscience. Philadelphia, USA: Lippincott Williams & Wilkins.

Bench, J., Kowal, Å. and Bamford, J., 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology, 13(3), pp.108-112.

Benson, K.B. ed., 1988. Audio engineering handbook. New York, USA: McGraw-Hill Companies.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M. and Rzeczkowski, C., 1984. Standardization of a test of speech perception in noise. Journal of Speech and Hearing Research, 27(1), p.32.

Blesser, B., 1969. Audio dynamic range compression for minimum perceived distortion. IEEE Transactions on Audio and Electroacoustics, 17(1), pp.22-32. New York, USA: IEEE.

Bohn, D.A., 2016. Pro Audio Reference. [online] Available at: <http://http://www.aes.org/par/guide/> [Accessed 3 Febuary 2011].

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D., 2000. A speech corpus for multitalker communications research. The Journal of the Acoustical Society of America, 107(2), pp.1065–1066.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spatial Vision, 10(4), pp.433–436.

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. Qualitative research in psychology, 3(2), pp.77-101.

Brown, S.W., 1997. Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. Perception & Psychophysics, 59(7), pp.1118–1140.

Clark, C. (2009) The Loudness Wars: Why Music Sounds Worse. In: NPR Music. [On-line] December 31. National Public Radio Inc., Washington D. C., USA. Available at:

<http://www.npr.org/2009/12/31/122114058/the-loudness-wars-why-music-sounds-worse> [Accessed 28 December 2015}.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J., 2011. Perception of emotional expression in musical performance. Journal of Experimental Psychology: Human Perception and Performance, 37(3), pp.921–934.

Bullock, J. and UCE Birmingham Conservatoire, 2007. Libxtract: a Lightweight Library for audio Feature Extraction. In: International Computer Music Association, The International Computer Music Conference (ICMC). Shanghai, China, 16th - 20th October 2017. Michigan, USA: Michigan Publishing

Bullock, J., 2008. Implementing audio feature extraction in live electronic music. Doctoral dissertation, PhD thesis, University of Birmingham.

Cabot, R. and Dennis, I., 2011. Loudness monitoring. [online] NewBay Media. Available at: <http://broadcastengineering.com/tests-and-measurements/BS1770-loudness-measurement-changing-0211/> [Accessed: 26 March 2011].

Cabot, R.C., 1999. Fundamentals of modern audio measurement. Journal of the Audio Engineering Society, 47 (9), pp.738-762.

A La Carta Televisión y Radio, Camerer, F., 2011. EBU_florian-camerer, Instituto RTVE [online] Available at: <http://www.rtve.es/alacarta/videos/television/ebu_florian-camerer/1034681/> [Accessed 04 September 2011].

Campbell, W., 2012. Compression: Investigations into the Loudness War. Cultures of the Digital Economy Conference, Anglia Ruskin University, Cambridge, UK. 27-28 March 2012.

Campbell, W., 2014. Strategies for reducing the effects of dynamic range compression. Faculty of Science & Technology: Pushing the Boundaries – 4th Annual Research and Scholarship Conference, Anglia Ruskin University, Cambridge, UK. 14 May 2014.

Campbell, W., Paterson, J. and Toulson, R., 2014. A quantitative evaluation of signal masking in summed and compressed audio. In: KES Transactions on Innovation in

Music, Special Edition - Innovation in Music 2013. Shoreham-by-Sea, UK: Future Technology Press, 1(1), pp.20-31.

Campbell, W., Toulson, R. and Paterson, J., 2010. The effect of dynamic range compression on the psychoacoustic quality and loudness of commercial music. 39th International Congress on Noise Control Engineering 2010 (Inter-Noise: Noise and Sustainability, 2010), Lisbon, Portugal, 13-16 June 2010. Red Hook, NY, USA: Curran Associates, Inc. pp.2580-2588.

Campbell, W., Paterson, J. and van der Linde, I., 2017. Listener preferences for alternative dynamic-range-compressed audio configurations. Journal of the Audio Engineering Society, 65(7/8), pp.540-551.

Cannam, C., Landone, C. and Sandler, M., 2010, Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 25-29 October 2010. New York, NY, USA: Association for Computing Machinery, pp. 1467-1468.

Case, A.U., 2011. Mix smart: Pro audio tips for your multitrack mix. Waltham, MA, USA: Focal Press.

Case, A., 2012. Sound FX: Unlocking the creative potential of recording studio effects. Boca Raton, FL, USA: CRC Press.

Cassidy, G.G. and MacDonald, R.A.R., 2010. The effects of music on time perception and performance of a driving game. Scandinavian Journal of Psychology, 51(6), pp.455-464.

Coryat, K., 2008. Guerrilla home recording: how to get great sound from any studio (no matter how weird or cheap your gear is). Milwaukee, WI, USA: Hal Leonard Corporation.

Croghan, N.B.H., Arehart, K.H. and Kates, J.M., 2012. Quality and loudness judgments for music subjected to compression limiting. The Journal of the Acoustical Society of America, 132(2), pp.1177-1188.

Cullari, S. and Semanchick, O., 1989. Music preferences and perception of loudness. Perceptual and Motor Skills, 68 (1), p.186.

Danckert, J.A. and Allman, A.A., 2005. Time flies when you're having fun: Temporal estimation and the experience of boredom. Brain and Cognition, 59 (3), pp.236-245.

De Man, B. and Reiss, J.D., 2013, October. A knowledge-engineered autonomous mixing system. Audio Engineering Society Convention 135. New York, NY, USA, 17-20 October 2013. New York: Audio Engineering Society.

De Man, B., Boerum, M., Leonard, B., King, R., Massenburg, G. and Reiss, J.D., 2015. Perceptual evaluation of music mixing practices. Audio Engineering Society Convention 138. Warsaw, Poland, 7-10 May 2015. New York: Audio Engineering Society.

De Man, B., Mora-Mcginity, M., Fazekas, G. and Reiss, J.D., 2014. An analysis and evaluation of audio features for multitrack music mixtures. Audio Engineering Society Convention 137. Los Angeles, CA, USA, 9-12 October 2014. New York: Audio Engineering Society.

De Man, B. and Reiss, J.D., 2015. Analysis of peer reviews in music production. In: Proceedings of the 2014 Art of Record Production Conference, Oslo, Norway, 4-6 December 2014. [online] Available at: <http://arpjournal.com/content/issue-10/> [Accessed 10 October 2016].

De Man, B., Leonard, B., King, R. and Reiss, J.D., 2014. An analysis and evaluation of audio features for multitrack music mixtures. In: Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014, Taipei, Taiwan, 27-31 October 2014. [online] Available at: <http://dblp.uni-trier.de/db/conf/ismir/ismir2014.html> [Accessed 10 October 2016].

Deruty, E. and Tardieu, D., 2014. About Dynamic Processing in Mainstream Music. Journal of the Audio Engineering Society, 62(1/2), pp.42-55.

Deutsch, D., 2013. Grouping mechanisms in music. The Psychology of Music. 3rd ed. Amsterdam, Netherlands: Elsevier. pp.183-248.

Devine, K., 2013. Imperfect sound forever: loudness wars, listening formations and the history of sound reproduction. Popular Music, 32(2), pp.159-176.

Dickason, V., 2006. Loudspeaker Design Cookbook. 7th ed. Peterburough, NH, USA: Audio Amateur Press.

Dixon, W.J. and Mood, A.M., 1946. The Statistical Sign Test. Journal of the American Statistical Association, 41(236), p.557.

Donahue, M., 2008. The Loudness War. Performer, [on-line] Available at: <http://performermag.com/Archives/loudness.php> [Accessed: 3 January 2011].

Doob, L. W.,1971. Patterning of time. London, England, UK: Yale University Press.

Droit-Volet, S. and Gil, S., 2009. The time–emotion paradox. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1525), pp.1943-1953.

Droit-Volet, S., Brunot, S. and Niedenthal, P., 2004. Brief Report: Perception of the duration of emotional events. Cognition and Emotion, 18(6), pp.849-858.

Droit–Volet, S., Tourret, S. and Wearden, J., 2004. Perception of the duration of auditory and visual stimuli in children and adults. The Quarterly Journal of Experimental Psychology Section A, 57(5), pp.797-818.

Doc, EBU Tech., 3341,“Loudness metering:‘EBU Mode’metering to Supplement Loudness Normalisation in Accordance with EBU R128”, Geneva, 2010.

Doc, EBU Tech., 3342,". Loudness Range: A descriptor to supplement loudness normalisation in accordance with EBU R, 128, Geneva, 2010.

Doc, EBU Tech., 3343,“Practical Guidelines for Production and Implementation in Accordance with EBU R128”, Geneva, 2011.

EBU, R., 2011. R 128-2011. Audio loudness normalisation & permitted maximum level, EBU-UER, Geneva.

EBU–Recommendation, R., 2011. Loudness normalisation and permitted maximum level of audio signals.

Ephraim, Y. and Malah, D., 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing, 33 (2), pp.443-445.

Epstein, M., 2013, June. Context effects in loudness. In Proceedings of Meetings on Acoustics ICA2013. Acoustical Society of America, 19(1), p. 050004.

Essling, C., Koenen, J. and Peukert, C., 2014. Loudness War: How Record Labels Compete for Consumer Attention. SSRN Electronic Journal.

Fletcher, H., and Munson, W.A., 1933. Loudness, Its Definition, Measurement, and Calculation. The Journal of the Acoustical Society of America, 5 (2), pp.82-108.

Fletcher, H., 1940. Auditory patterns. Reviews of modern physics, 12 (1), pp.47.

Flexer, C., 2002. Rationale and use of sound field systems: An update. The Hearing Journal, 55 (8), pp.10-18.

Fraisse, P., 1978. Time and rhythm perception. In: Handbook of perception, Vol. 8: Perceptual coding, ed. E. C. Carterette & M. P. Friedman, pp.203-254. New York: Academic Press.

Fuchs, P. ed., 2010. Oxford Handbook of Auditory Science: The Ear. Oxford, England, UK: Oxford University Press.

Gazzaniga, M., Ivry, R. and Mangun, G., 2002. Cognitive neuroscience: the biology of the mind. 2nd ed. New York: WW Norton & Company. Inc.

Giannoulis, D., Massberg, M. and Reiss, J.D., 2012. Digital dynamic range compressor design - a tutorial and analysis. Journal of the Audio Engineering Society, 60(6), pp.399-408.

Glasberg, B.R. and Moore, B.C., 1990. Derivation of auditory filter shapes from notched-noise data. Hearing research, 47 (1), pp.103-138.

Glasberg, B.R. and Moore, B.C.J., 2002. A Model of Loudness Applicable to Time-Varying Sounds. Journal of the Audio Engineering Society, 50 (5), pp.331-342.

Gorlow, S. and Reiss, J.D., 2013. Model-based inversion of dynamic range compression. IEEE Transactions on Audio, Speech, and Language Processing, 21(7), pp.1434-1444.

Grondin, S., 2008. Methods for studying psychological time. Psychology of time, pp.51-74.

Hellman, R.P., 1985. Perceived Magnitude of Two-Tone-Noise Complexes: Loudness, Annoyance, and Noisiness. Journal of the Audio Engineering Society, 77(4), pp.1497-1504.

Helmholtz, H., 1954. On the sensations of tone, 1877. Translated from German by A.J. Ellis. Dover, NY, USA, pp.390-394.

Hjortkjær, J. and Walther-Hansen, M., 2014. Perceptual effects of dynamic range compression in popular music recordings. Journal of the Audio Engineering Society, 62(1/2), pp.37–41.

Hoover, D.M. and Cullari, S., 1992. Perception of loudness and musical preference: comparison of musicians and non-musicians. Perceptual and motor skills, 74 (3 Pt 2), pp.1149-1150.

Hoeg, W., Christensen, L., and Walker, R., 1997. Subjective assessment of audio quality-the means and methods in the EBU. EBU Technical Review, pp.40-50.

Irwin, J., 2006. Basic Anatomy and physiology of the Ear. Infection and Hearing Impairment, pp.8-13.

ISO, 226: 2003. Acoustics—Normal equal loudness-level contours. International Organization for Standardization, Geneva, Switzerland.

Izhaki, R., 2013. Mixing audio: concepts, practices, and tools. Abingdon, England, UK: Taylor & Francis.

Jacobs, L., 2012. The Innovation of Re-Recording in the Hollywood Studios. Film History: An International Journal, 24 (1), pp.5-34.

Jones, S., 2005. The Big Squeeze, Mastering Engineers Debate on the Loudness War. [on-line] Available at: <http://mixonline.com/mag/audio_big_squeeze/>. [Accessed: 11 October 2009].

Juslin, P.N. and Laukka, P., 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129(5), pp.770–814.

Kahneman, D., 1973. Attention and effort (Vol. 1063). Englewood Cliffs, NJ, USA: Prentice-Hall..

Kamenetsky, S.B., Hill, D.S. and Trehub, S.E., 1997. Effect of tempo and dynamics on the perception of emotion in music. Psychology of Music, 25(2), 149-160.

Kantor-Martynuska, J., 2009. The listener's temperament and perceived tempo and loudness of music. European Journal of Personality, 23 (8), pp.655-673.

Kates, J.M. and Arehart, K.H., 2010. The hearing-aid speech quality index (HASQI). Journal of the Audio Engineering Society, 58 (5), pp.363-381.

Kates, J.M. and Arehart, K.H., 2014. The hearing-aid speech quality index (HASQI) version 2. Journal of the Audio Engineering Society, 62 (3), pp.99-117.

Kates, J.M., 2005. Principles of digital dynamic-range compression. Trends in amplification, 9 (2), pp.45-76.

Katz, B., 2007. Mastering Audio, the Art and Science. 2nd ed. Waltham, MA, USA: Focal Press.

Kellaris, J.J. and Rice, R.C., 1993. The influence of tempo, loudness, and gender of listener on responses to music. Psychology & Marketing, 10 (1), pp.15-29.

Kirchberger, M. and Russo, F.A., 2016. Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. Trends in hearing, 20, p.2331216516630549.

Kirk, R.E., 1956. Learning, a major factor influencing preferences for high-fidelity reproducing systems. The Journal of the Acoustical Society of America, 28(6), pp.1113–1116.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R. and Broussard, C., 2007. What's new in Psychtoolbox-3. Perception, 36(14), p.1.

Kramer, S.E., Kapteyn, T.S. and Houtgast, T., 2006. Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work: Desempeño laboral: Comparación de empleados con audición normal o alterada usando el Listado Amsterdam para Audición y Trabajo. International journal of audiology, 45 (9), pp.503-512.

Lawless, H.T., 2013. Quantitative sensory analysis: psychophysics, models, and intelligent design. Hoboken, NJ, USA: John Wiley & Sons.

Le Brun, M., 1979. Digital waveshaping synthesis. Journal of the Audio Engineering Society, 27 (4), pp.250-266.

Levine, R., 2007. The Death of High Fidelity. [on-line] Available at: <http://www.rollingstone.com/news/story/17777619/the_death_of_high_fidelity/1>. [Accessed: 7 November 2009].

Levitin, D.J., 2011. This is your brain on music: Understanding a human obsession. London, England, UK: Atlantic Books Ltd.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychology. 140, pp.1–55.

Ma, Z., De Man, B., Pestana, P., Black, D. and Reisss, J., 2015. Intelligent Multitrack Dynamic Range Compression. Journal of the Audio Engineering Society, 63(6), pp.412-426.

Macmillan, N.A. and Creelman, D.C., 2004. Detection Theory: A User’s Guide. Abingdon, England, UK: Routledge.

Maddams, J.A., Finn, S. and Reiss, J.D., 2012. An autonomous method for multi-track dynamic range compression. In: Proceedings of the 15th International Conference on Digital Audio Effects (DAFx12). York, England, UK, 17-21 September 2012. [online] Available at: <https://www.dafx12.york.ac.uk/proceedings.html> [Accessed 10 October 2016].

"Maempel, H.J. and Gawlik, F., 2009, May. The Influence of Sound Processing on Listeners' Program Choice in Radio Broadcasting. In: Audio Engineering Society Convention 126. Munich, Germany, 7–10 May 2009. New

York, NY, USA: Audio Engineering Society."

Mahkonen, K., Eronen, A., Virtanen, T., Helander, E., Popa, V., Leppänen, J., Curcio, I.D., Čierny, M., Wang, H. and Wichman, R., 2013. Music dereverberation by spectral linear prediction in live recordings. In: Procedings of the 16th International Conference on Digital Audio Effects (DAFx-13). Maynooth, Republic of Ireland: 2-5 September 2013. [online] Available at: <http://dafx13.nuim.ie/> [Accessed 25 June 2015].

Hine, R. and Martin, E. eds., 2015. A dictionary of biology. Oxford Quick Reference. Oxford, England, UK: Oxford University Press.

Marven, C. and Ewers, G., 1996. A simple approach to digital signal processing. New York, NY, USA: John Wiley & Sons, Inc.

McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore, D.R., Barry, J.G. and Amitay, S., 2014. Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. International journal of audiology, 53(7), 433-440.

McGregor, K.K., Berns, A.J., Owen, A.J., Michels, S.A., Duff, D., Bahnsen, A.J. and Lloyd, M., 2012. Associations between syntax and the lexicon among children with or without ASD and language impairment. Journal of Autism and Developmental Disorders, 42 (1), pp.35-47.

Death Magnetic, 2008. [Sound recording] Performed by Metallica. U.S.A: Vertigo Records.

Metallicabb.com, 2008. I got a response from Ted [Anything Metallica]. 12 September. Available at: <http://www.metallicabb.com/index.php?showtopic=85317> [Accessed: 17 July 2012].

Milner, G., 2009. Perfecting sound forever: an aural history of recorded music. London, England, UK: Macmillan.

Moore, A., 2012. All Buttons In: An investigation into the use of the 1176 FET compressor in popular music production. In: Proceedings of the 2014 Art of Record Production Conference, Oslo, Norway, 4-6 December 2014. [online] Available at: <http://arpjournal.com/content/issue-10/> [Accessed 10 October 2016].

Moore, B.C. and Tan, C., 2003. Perceived naturalness of spectrally distorted speech and music. The Journal of the Acoustical Society of America, 114(1), pp.408-419.

Moore, B.C., 2013. An introduction to the psychology of hearing. 4th ed. Leiden, Netherlands: Brill.

Moore, B.C., Tan, C.-T., Zacharov, N. and Mattila, V.-V., 2004. Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion. Journal of the Audio Engineering Society, 52(12), pp.1228–1244.

Moore, D.R., Fuchs, P.A., Rees, A., Palmer, A.R. and Plack, C., 2010. Oxford Handbook of Auditory Science: Hearing. Oxford, England, UK: Oxford University Press.

Moylan, W., 2007. Understanding and crafting the mix: The art of recording. Waltham, MA, USA: Focal Press.

Mynett, M., 2017. Metal Music Manual: Producing, Engineering, Mixing, and Mastering Contemporary Heavy Music. Abingdon, England, UK: Taylor & Francis.

Neuhoff, J.G., McBeath, M.K. and Wanzie, W.C., 1999. Dynamic frequency change influences loudness perception: A central, analytic process. Journal of experimental psychology, 25 (4), pp.1050-1059.

Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman, S. and Levitt, H., 1998. The effect of compression ratio and release time on the categorical rating of sound quality. The Journal of the Acoustical Society of America, 103(5), pp.2273–2281.

Nichols, A.L. and Maner, J.K., 2008. The good-subject effect: Investigating participant demand characteristics. The Journal of general psychology, 135(2), pp.151–166.

Nielsen, S.H. and Lund, T., 2000. 0 dB FS+ Levels in Digital Mastering. In: Audio Engineering Society Convention 109. Los Angeles, CA, USA, 22-25 September 2000. New York, NY, USA: Audio Engineering Society.

Nielsen, S.H. and Lund, T., 2003. Overload in signal conversion. In Audio Engineering Society Conference. In: 23rd International Audio Engineering Society Conference: Signal Processing in Audio Recording and Reproduction. Copenhagen, Denmark, 23-25 May 2003. New York, NY, USA: Audio Engineering Society.

Nisbett, A., 1970. The Technique of the Sound Studio: Radio, Record Production, Television, and Film. 2nd ed. Norwalk, CT, USA: Hastings House

North, A. and Hargreaves, D., 1999. Can Music Move People? The Effects of Musical Complexity and Silence on Waiting Time. Environment and Behaviour, 31 (1), pp.136-149.

Noulhiane, M., Mella, N., Samson, S., Ragot, R. and Pouthas, V., 2007. How emotional auditory stimuli modulate time perception. Emotion, 7 (4), pp.697.

Oehler, M., Reuter, C. and Czedik-Eysenberg, I., 2015. Dynamics and Low-Frequency Ratio in Popular Music Recordings since 1965. In: Audio Engineering Society 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television, and the Internet. Hollywood, CA, USA, 6-8 March 2015. New York, NY, USA: Audio Engineering Society.

Olive, S., 2011. Some new evidence that teenagers may prefer accurate sound reproduction. In: Audio Engineering Society Convention 131. New York, NY, USA, 20-23 October, 2011. New York, NY, USA: Audio Engineering Society.

Olive, S., 2012. Some new evidence that teenagers may prefer accurate sound reproduction. In: Audio Engineering Society Convention 132. Budapest, Hungary, 26-29 April, 2012. New York, NY, USA: Audio Engineering Society.

Olson, H.F., 1947. Frequency range preference for speech and music. The Journal of the Acoustical Society of America, 19(4), pp.549–555.

Open Congress, 2008. H.R.6209: Commercial Advertisement Loudness Mitigation Act - U.S. Congress - OpenCongress. [on-line] Available at: <http://www.opencongress.org/bill/110-h6209/show> [Accessed: 9 March 2011].

Owsinski, B., 2013. The mixing engineer's handbook. Scarborough, ON, Canada: Nelson Education.

Packwood, W.T. and Parker, C.A., 1973. A method for rating counsellor social reinforcement and persuasion. Journal of Counselling Psychology, 20(1), pp.38-43.

Packwood, W.T., 1974. Loudness as a variable in persuasion. Journal of Counselling Psychology, 21 (1), pp.1-2.

Pearson, K., 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5, 50(302), pp.157–175.

Peeters, G., Giordano, B.L., Susini, P., Misdariis, N. and McAdams, S., 2011. The timbre toolbox: Extracting audio descriptors from musical signals. The Journal of the Acoustical Society of America, 130(5), pp.2902-2916.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision. 10(4), pp.437–442.

Pestana, P. and Reiss, J., 2014. Intelligent Audio Production Strategies Informed by Best Practices. Journal of the Audio Engineering Society. 63(6), pp.412–426.

Pestana, P.D., Reiss, J.D. and Barbosa, A., 2013. Loudness measurement of multitrack audio content using modifications of ITU-R BS. 1770. In: Audio Engineering Society Convention 134. Rome, Italy, 4-7 May, 2013. New York, NY, USA: Audio Engineering Society.

Pestana, P.D.L.G., 2013. Automatic mixing systems using adaptive digital audio effects. PhD Universidade Católica Portuguesa.

Pierce, J.R., 1983. The science of musical sound. New York, NY, USA: Scientific American Library.

Plack, C.J. and Moore, D.R., 2010. The oxford handbook of auditory science: hearing. Oxford, England, UK: Oxford University Press.

Plenge, G., Spikofski, G. and Theile, G., eds. 1985. Individual Choice of the Dynamic Range for Radio Programs. Audio Engineering Society Convention 79. New York, NY, USA, 12-16 October, 1985. New York, NY, USA: Audio Engineering Society.

Price, J.L., 2005. Free will versus survival: brain systems that underlie intrinsic constraints on behaviour. Journal of Comparative Neurology, 493(1), pp.132-139.

Rabbitt, P., 1991. Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. Acta Oto-Laryngologica, 111 (sup 476), pp.167-176.

Rabbitt, P.M., 1968. Channel-capacity, intelligibility, and immediate memory. The Quarterly journal of experimental psychology, 20 (3), pp.241-248.

ITU Recommendation, 1990. BS. 562-3, Subjective Assessment of Sound Quality, Geneva: International Telecommunication Union.

Reiss, J., 2016. A Meta-Analysis of High Resolution Audio Perceptual Evaluation. Journal of the Audio Engineering Society, 64(6), pp.364–379.

Law, J. and Rennie, R. eds., 2015. A dictionary of physics. 7th ed. Oxford, England, UK: Oxford University Press.

Riley, K.G. and McGregor, K.K., 2012. Noise hampers children's expressive word learning. Language, speech, and hearing services in schools, 43 (3), pp.325-337.

Roads, C., 1979. A Tutorial on Non-Linear Distortion or Waveshaping Synthesis. Computer Music Journal, 3(2), pp.29–34.

Roberts, D. ed., 2002. Signals and perception: the fundamentals of human sensation. Basingstoke, England, UK: Palgrave.

Robinson, C.E. and Huntington, D.A., 1973. The Intelligibility of Speech Processed by Delayed Long-Time-Averaged Compression Amplification. The Journal of the Acoustical Society of America, 54, p.314.

Robinson, D.W. and Dadson, R.S., 1956. A re-determination of the equal-loudness relations for pure tones. British Journal of Applied Physics, 7(5), pp.166-181.

Ronan, D., De Man, B., Gunes, H. and Reiss, J.D., 2015. The Impact of Subgrouping Practices on the Perception of Multitrack Music Mixes. In: Audio Engineering Society

Convention 139. New York, NY, USA, 29 October - 1 November, 2015. New York, NY, USA: Audio Engineering Society.

Ronan, D.M., Gunes, H. and Reiss, J.D., 2017. Analysis of the subgrouping practices of professional mix engineers. In: Audio Engineering Society Convention 142. Berlin, Germany, 20-23 May, 2017. New York, NY, USA: Audio Engineering Society.

Ronan, M., Sazdov, R. and Ward, N., 2014. Factors influencing listener preference for dynamic range compression. In: Audio Engineering Society Convention 137, Los Angeles, CA, USA, 9-12 October, 2014. New York, NY, USA: Audio Engineering Society.

Davis, M.F., , 2007. Audio and Electroacoustics. In: Rossing, T.D. ed., Handbook of Acoustics, New York, NY, USA: Springer, pp. 743–781.

Rudner, M., Lunner, T., Behrens, T., Thorén, E.S. and Rönnberg, J., 2012. Working memory capacity may influence perceived effort during aided speech recognition in noise. Journal of the American Academy of Audiology, 23(8), pp.577-589.

Rumsey, F., 2008. If it's Loud does that mean It's Bad? Two workshops on broadcast issues from the 123rd AES convention. Journal of the Audio Engineering Society, 56(6), pp.493-498.

Sackett, A. M., Meyvis, T., Nelson, L. D., Converse, B. A., & Sackett, A. L., 2010. You're having fun when time flies: The hedonic consequences of subjective time progression. Psychological Science, 21(1), pp.111-117.

Sarampalis, A., Kalluri, S., Edwards, B. and Hafter, E., 2009. Objective measures of listening effort: Effects of background noise and noise reduction. Journal of Speech, Language, and Hearing Research, 52(5), pp.1230-1240.

Scharf, B., 2001. Sequential effects in loudness. State College, PA, USA: CiteSeerX.

Schoenherr, S.E., 2004. Tom Stockham and Digital Audio Recording [on-line] Available at: <http://www.aes.org/aescms.cfm?aescmspath=aeshc/docs/recording.technology.history/stockham.html&rndx=566146> [Accessed 18 July 2016].

Schubert, E., 2004. Modelling perceived emotion with continuous musical features. Music perception, 21(4), pp.561-585.

Self, D., Duncan, B., Sinclair, I., Brice, R., Hood, J.L., Singmin, A., Davis, D., Patronis, E. and Watkinson, J., 2009. Audio engineering: Know it all. Oxford, UK: Newnes.

ITU Recommendation, 2011. BS.1770-3. Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level, Geneva: International Telecommunications Union.

ITU Recommendation, 2012. BS.1770-4. Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level, Geneva: International Telecommunications Union.

Serrà, J., Corral, Á., Boguñá, M., Haro, M. and Arcos, J.L., 2012. Measuring the evolution of contemporary western popular music. Scientific reports, 2, p.521.

Shelvock, M., 2012. Audio Mastering as Musical Practice, Doctoral dissertation, PhD thesis, The University of Western Ontario.

Shepherd, I., 2008. Metallica "Death Magnetic" sounds better in Guitar Hero - Mastering Media Blog. [blog] 15 September. Available at: <http://mastering-media.blogspot.com/2008/09/metallica-death-magnetic-sounds-better.html> [Accessed: 15 April 2011].

Sherwin, A., 2007. Why Music Really Is Getting Louder. Times (London) [on-line] Available at: <http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/music/article1878724.ece>.

Shuttleworth, T., ed. 2013. Audio Engineering Society Convention 135, New York, NY, USA, 17-20 October, 2013. New York, NY, USA: Audio Engineering Society.

Smith, S.E., 1999. Dancing in the street: Motown and the cultural politics of Detroit. Cambridge, MA, USE: Harvard University Press.

Solberg, R.T., 2014. ‘Waiting for the Bass to Drop’: Correlations Between Intense Emotional Experiences and Production Techniques in Build-up and Drop Sections of Electronic Dance Music. Dancecult, 6(1), pp.61-82.

Sterne, J., 2006. The death and life of digital audio. Interdisciplinary Science Reviews, 31(4), pp.338-348.

Stevens, S.S., 1957. On the psychophysical law. Psychological review, 64 (3), pp.153.

Stone, M.A., and Moore, B.C.J., 2003. Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task. The Journal of the Acoustical Society of America, 114(2), 1023-1034.

Stone, M.A., Moore, B.C.J., Füllgrabe, C. and Hinton, A.C., 2009. Multichannel Fast-Acting Dynamic Range Compression Hinders Performance by Young Normal-Hearing Listeners in A Two Two-Talker Separation Task. Journal of the Audio Engineering Society, 57(7/8), pp.532–546.

Takeshima, H., Suzuki, Y., Kumagai, M., Sone, T., Fujimori, T. and Miura, H., 1994. Threshold of hearing for pure tone under free-field listening conditions. Journal of the Acoustical Society of Japan (E), 15(3), pp.159-169.

Tan, C.-T. and Moore, B.C., 2004. Comparison of two forms of fast-acting compression using physical and subjective measures. In: Proceedings of the 18th International Congress on Acoustics, Kyoto, Japan, 4-9 April, 2004.

Commercial Advertisement Loudness Mitigation (CALM) Act Summary of Best Practices. (2012). NAB TV Technology Committee. NAB Labs. p.2. [on-line] Available at: <http://www.opencongress.org/bill/110-h6209/text> [Accessed: 27 May 2011].

Tischmeyer, F., 2009. Pleasureize Music Foundation. [on-line] Available at: <http://dynamicrange.de/en/our-aim> [Accessed: 9 March 2011].

Toulson, E.R. and Campbell, W., 2009. The life and death of dynamic range: who decides how loud? The Art of Record Production Conference, Cardiff, Wales, UK, 13-15 November, 2009.

Toulson, E.R., Paterson, J.L. and Campbell, W., 2014. Evaluating harmonic and intermodulation distortion of mixed signals processed with dynamic range compression. In: KES Transactions on Innovation in Music, Special Edition - Innovation in Music 2013. Shoreham-by-Sea, UK: Future Technology Press, 1(1), pp.224–246.

Tulving, E., 2002. Episodic memory: From mind to brain. Annual Review of Psychology, 53(1), pp.1-25.

van Buuren, R.A., Festen, J.M. and Houtgast, T., 1999. Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. The Journal of the Acoustical Society of America, 105(5), pp.2903–2913.

Vickers, E., 2010. The loudness war: Background, speculation, and recommendations. Audio Engineering Society Convention 129, San Francisco, CA, USA, 4-7 November, 2010. New York, NY, USA: Audio Engineering Society.

Viney, D., 2008. The obsession with compression. Master of Music Technology Thesis, London College of Music, Thames Valley University.

Ward, J. and Campbell, W., 2010. Loss of our musical heritage: The rise of the digital re-master. The Art of Record Production Conference, Leeds Metropolitan University, Leeds, UK. 3-5 December 2010.

Watkinson, J., 2002. An introduction to digital audio. Abingdon, England, UK: Taylor & Francis.

Weiss, B., 2010. Rock on with vintage jukeboxes. [on-line] Antique Week. Available at: <http://www.antiqueweek.com/Article.asp?newsid=1796> [Accessed: 3 January 2016].

Wendl, M. and Lee, H., 2014. The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor. Audio Engineering Society Convention 136. Berlin, Germany, 26-29 April, 2014. New York, NY, USA: Audio Engineering Society.

Yoshida, J., Hasegawa, H. and Kasuga, M., 2006. Previous sound effects on loudness in paired comparison experiments. Acoustical science and technology, 27(3), pp.147-153.

Zielinski, S., Rumsey, F. and Bech, S., 2008. On Some Biases Encountered in Modern Audio Quality Listening Tests–A Review. Journal of the Audio Engineering Society, 56(6), pp.427–451.

Zölzer, U. ed., 2011. DAFX: digital audio effects. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons.